دانشکده علوم ریاضی
گروه آمار

پایان‌نامه کارشناسی ارشد

همگرایی ارگودیک یکنواخت و هندسی الگوریتم‌های مونت‌کارلوی زنجیر مارکوفی مولفه به مولفه

نگارنده: فاطمه خدابخشی پالندی

استاد راهنما

دکتر نگار اقبال

استاد مشاور

دکتر حسین باغیشنی

شهریور ۱۳۹۵

تقدیم با بوسه به دستان استوارترین تکیه گاهانم، پدر فداکار و مادر عزیزتر از جانم...

که هر آن چه آموختم در مکتب عشق شما آموختم و هرچه بکوشم قطره ای از دریای بی کران مهربانی تان را سپاس نتوانم بگویم. امروز هستی ام به امید شماست و فردا کلید باغ بهشتم رضای شما...

باشد که حاصل تلاشم نسیم گونه، غبار خستگی تان را بزداید وگرنه به پاس زحمات شما، جان را سزاوار است.

تقدیم به پناه خستگی هایم و امید بودنم، همسر مهربانم...

او که اسوه صبر و تحمل بوده و مشکلات مسیر را برایم تسهیل نمود.

تقدیم به خواهرها و برادرهایم، همراهان همیشگی زندگیم...

آنان که وجودشان باعث آرامش و اطمینان است.

پروردگارا، حسن عاقبت، سلامت و سعادت را برای آنان مقدر بفرما...

بارالها...
از کوی تو بیرون نشود...
پای خیالم،
نکند فرق به حالم...
چه برانی، چه بخوانی...
چه به اوجم برسانی...
چه به خاکم بنشانی،
نه من آنم که بر نجم...
نه تو آنی که برانی...
نه من آنم که ز فیض نکهت چشم پوشم،
نه تو آنی که گدا را ننوازی به نگاهی...
در اگر باز نگردد...
نروم باز به جایی،
پشت دیوار نشینم چو گدا بر سر راهی...
کس به غیر از تو نخواهم...
چه بخواهی چه نخواهی،
باز کن در که جز این خانه مرا نیست پناهی.

خواجه عبدالله انصاری

شکر شایان ایزد منان که توفیق رارفیق راهم ساخت تا این پایان نامه را به پایان برسانم. اکنون بر خود واجب می دانم که ارج نهم بزرگانی راکه آورده های شمع وجودشان رو شنگر راهم است.

سپاس فراوان خود راتقدیم می دارم به محضر استاد ارجمندم سرکار خانم دکتر اقبال که از ابتدای راه در طی انجام این پایان نامه با راهنمایی های خود مرادر نگارش این اثر یاری نمودند. تقدیر و تشکر خود را از استاد مشاور جناب آقای دکتر باغیشنی صمیمانه ابراز می دارم، چراکه بدون راهنمایی های ایشان به اتمام رساندن این پایان نامه بسیار مشکل می نمود. از اساتید محترم جناب آقای دکتر آرشی و سرکار خانم دکتر دستنج که زحمت داوری این اثر رابر عهده گرفته اند کمال تشکر را دارم.

در پایان، تشکر خالصانه خدمت همه کسانی که به نوعی مرادر به انجام رساندن این مهم یاری نموده اند.

نگارنده: فاطمه خدابخشی پالندی

شهریور ۱۳۹۵

تعهدنامه

اینجانب **نگارنده: فاطمه خدابخشی پالندی** دانشجوی کارشناسی ارشد رشته **آمار دانشکده علوم ریاضی** دانشگاه صنعتی شاهرود، نویسنده پایان‌نامه با عنوان **همگرایی ارگودیک یکنواخت و هندسی الگوریتم‌های مونت‌کارلوی زنجیر مارکوفی مولفه به مولفه** ، تحت راهنمایی **دکتر نگار اقبال** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود متعلق دارد، و مقالات مستخرج با نام " دانشگاه صنعتی شاهرود " یا " Shahrood University of Technology " به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده) شده است، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

نگارنده: فاطمه خدابخشی پالندی

شهریور ۱۳۹۵

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

در دیدگاه استنباط بیزی، تمامی استنباط‌ها مبتنی بر توزیع پسین به‌دست می‌آیند. در اغلب موقعیت‌های کاربردی توزیع پسین صورت بسته‌ای ندارد و باید به شیوه‌ای آن‌را تقریب زد. در روش‌های مونت‌کارلوی زنجیر مارکوف، معمول است که به جای به‌روز رسانی هم‌زمان همه متغیرها، در هر مرحله فقط یک متغیر (یا یک بلوک‌های متغیر) به‌روز رسانی می‌شود که به آن به‌روز رسانی مولفه به مولفه می‌گویند. الگوریتم‌های نمونه‌گیر گیبز و متروپولیس-هستینگز درون گیبز از این جمله هستند. پس از به‌روز رسانی به روش مولفه به مولفه همه متغیرها (یا بلوک‌های متغیر)، استراتژی‌های مختلفی برای ترکیب آن‌ها، از جمله اتصال همگی یا انتخاب یک دنباله تصادفی، به عنوان حالت زنجیر توام همه متغیرها قابل به‌کارگیری هستند. در کنار مزیت های روش‌های به‌روز رسانی مولفه به مولفه، مطالعه و پژوهش بر روی ویژگی‌های نظری همگرایی زنجیرهای مارکوف حاصل از آن‌ها به توزیع مانای خود، که همان توزیع پسین مورد نظر می‌باشد، به ندرت مورد توجه قرار گرفته است. ما شرایطی که تحت آن برخی زنجیرهای مارکوف مولفه به مولفه به یک توزیع ثابت با نرخ هندسی همگرا می‌شوند را مطالعه می‌کنیم و توجه ویژه‌ای به ارتباط بین نرخ‌های همگرایی و استراتژی‌های مولفه به مولفه گوناگون داریم. در نهایت نتایج را، برای دو مثال واقعی که شامل، یک مدل سلسله مراتبی خطی آمیخته و برآورد ماکسیمم درستنمایی برای مدل آمیخته، نشان می‌دهیم.

کلمات کلیدی:
ارگودیک هندسی، ارگودیک یکنواخت، زنجیر مارکوف، مونت‌کارلو، نمونه‌گیر گیبز، متروپولیس-هستینگز درون گیبز، اسکن تصادفی، نرخ همگرایی

پیش گفتار

در دیدگاه استنباط بیزی، تمامی استنباط‌ها مبتنی بر توزیع پسین به‌دست می‌آیند. در اغلب موقعیت‌های کاربردی توزیع پسین صورت بسته‌ای ندارد و باید به شیوه‌ای آن‌را تقریب زد. یک رده از روش‌ها برای تقریب توزیع پسین، روش‌های نمونه‌گیری MCMC، به‌ویژه الگوریتم‌های متروپولیس-هستینگز و نمونه‌گیر گیبز، می‌باشد. این روش‌ها، بنابر زنجیرهای مارکوف طوری طراحی می‌شوند که با افزایش تعداد تکرارهای نمونه‌گیری یا به عبارتی با به‌روز رسانی حالت زنجیر از جایی به بعد، حالت‌های زنجیر حاصل به عنوان نمونه هایی از توزیع مطلوب، یعنی توزیع پسین، در نظر گرفته می‌شوند.

در مسایل واقعی، بعد متغیرهای پسین بالاست و انتخاب توزیع پیشنهادی می‌تواند چالش برانگیز باشد و در روش‌های MCMC، معمول است که به جای به‌روز رسانی هم‌زمان همه متغیرها، در هر مرحله فقط یک متغیر (یا یک بلوک‌های متغیر) به‌روز رسانی می‌شود که به آن به‌روز رسانی مولفه به مولفه اطلاق می‌شود (نیل و رابرتز، ۲۰۰۶). الگوریتم‌های نمونه‌گیر گیبز و متروپولیس-هستینگز درون گیبز از این جمله هستند. پس از به‌روز رسانی به روش مولفه به مولفه همه متغیرها (یا بلوک‌های متغیر)، استراتژی‌های مختلفی برای ترکیب آن‌ها، از جمله اتصال همگی یا انتخاب یک دنباله تصادفی، به عنوان حالت زنجیر توام همه متغیرها قابل به‌کارگیری هستند (فورت و همکاران، ۲۰۰۳). این روش‌ها و استراتژی‌های ترکیب کردن، باعث اجرای ساده‌تر و عملکرد به‌تر الگوریتم MCMC نسبت به روش‌های به‌روز رسانی هم‌زمان همه متغیرها می‌شوند. در کنار مزیت های روش‌های به‌روز رسانی مولفه به مولفه، مطالعه و پژوهش بر روی ویژگی‌های نظری همگرایی زنجیرهای مارکوف حاصل از آن‌ها به توزیع مانای خود، که همان توزیع پسین مورد نظر می‌باشد، به ندرت مورد توجه قرار گرفته است.

در مورد همگرایی یکنواخت و هندسی به‌صورت‌های مختلف زنجیرهای متروپولیس-هستینگز، زمانی‌که همه متغیرها هم‌زمان به‌روز می‌شوند، پژوهش‌های زیادی انجام شده‌اند. به عنوان چند نمونه می‌توان به منگرسن و تویدی (۱۹۹۶)، کریستنسن و همکاران (۲۰۰۹) و جانسون و گیر (۲۰۱۲) اشاره کرد. اما هیچ‌کدام از این نتایج برای زنجیرهای مولفه به مولفه متروپولیس-هستینگز به‌کار گرفته نشده‌اند. تنها روش مولفه به مولفه‌ای که کمی مورد توجه قرار گرفته است نمونه‌گیر گیبز است.

اثبات همگرایی زنجیرهای MCMC مولفه به مولفه (به همراه استراتژی‌های مختلف ترکیب) با یک نرخ همگرایی هندسی می‌تواند بسیار مهم باشد زیرا برای کاربران آن‌ها این اطمینان را حاصل می‌کند که می‌توانند از نمونه‌های تولید شده از چنین زنجیرهایی با عنوان نمونه‌های با کیفیت نزدیک به نمونه‌های i.i.d. استفاده کنند (منگال و جونز، ۲۰۱۱). به همین دلیل در این پایان‌نامه بر آن شدیم تا شرایطی که تحت آن‌ها برخی از زنجیرهای مارکوف مولفه به مولفه با یک نرخ هندسی به توزیع مانای خود همگرا می‌شوند را نمایش دهیم و همچنین روابط بین نرخ‌های همگرایی حاصل از استراتژی‌های مختلف ترکیب را بررسی کنیم. مطالب پایان‌نامه، به اختصار، شامل موارد زیر می‌باشد:

- فصل اول، تعاریف و مفاهیم مورد نیاز برای ورود به موضوع این پایان‌نامه را مطرح می‌کنیم.
- در فصل دوم، بر روی برخی نماد، فضای حالت و توسعه یک چارچوب کلی برای به‌روز رسانی مولفه

به مولفه تمرکز می‌کنیم.

- در فصل سوم، نرخ‌های همگرایی روش‌های مولفه به مولفه و به‌خصوص ارتباط بین نرخ همگرایی نمونه‌های اسکن قطعی با دنباله‌ای از اسکن تصادفی و روش‌های اسکن تصادفی مطالعه می‌کنیم و همچنین شرایط برای ایجاد ارگودیک یکنواخت نسخه هایی از الگوریتم متروپولیس-هستینگز با توزیع پیشنهادی در حالت مستقل به‌وسیله روش مولفه به مولفه بررسی می‌کنیم.
- در فصل چهارم، به منظور ارزیابی به‌روز رسانی به روش مولفه به مولفه و مقایسه آن با روش به‌روز رسانی هم‌زمان همه متغیرها، برای دو مثال، نمونه‌گیری گیبز برای مدل خطی آمیخته و برآورد ماکسیمم درستنمایی برای مدل‌های آمیخته، مطالعه شبیه‌سازی انجام شده است. نتایج، حاکی از کارایی مطلوب‌تر روش به‌روز رسانی مولفه به مولفه، نسبت به روش به‌روز رسانی هم‌زمان همه متغیرها می‌باشد.

فهرست نشانه‌ها و نمادها

Y متغیر تصادفی
$\mathbf{y}$ بردار تصادفی
$\pi(\theta|y)$ توزیع پسین θ به شرط y
$\pi(\theta)$ توریع پیشین θ
χ فضای حالت
Θ فضای پارامتر
$L(\theta|y)$ تابع درستنمایی
$\mathcal{B}$ سیگما جبر بورل
$\rightarrow$ میل می‌کند
$\xrightarrow{\mathrm{d}}$ همگرایی در توزیع
Φ زنجیر مارکوف
MCMC مونت‌کارلوی زنجیر مارکوف
P_{comp} هسته ترکیب
P_{mix} هسته آمیختگی
P_{RS} هسته اسکن تصادفی
P_C هسته ترکیب
P_{RQ} هسته دنباله اسکن تصادفی
P_{GS} هسته نمونه‌گیر گیبز
P_{HC} هسته ترکیب متروپولیس-هستینگز درون گیبز
P_{RSH} هسته اسکن تصادفی متروپولیس-هستینگز درون گیبز
P_{CIS} هسته نمونه ترکیب مستقل
P_{RQIS} هسته نمونه دنباله تصادفی مستقل
P_{RSIS} هسته نمونه اسکن تصادفی مستقل
h_{RS} چگالی انتقال مارکوف اسکن تصادفی
h_C چگالی انتقال مارکوف ترکیب
h_{RQ} چگالی انتقال مارکوف دنباله اسکن تصادفی
h_{HC} چگالی انتقال مارکوف ترکیب متروپولیس-هستینگز درون گیبز
$\|\cdot\|$ تابع نرم

فهرست مطالب

لیست تصاویر

لیست جداول

فصل ۱

تعاریف و مفاهیم اولیه

در این فصل، روش‌ها، مفاهیم و تعاریف لازم برای ورود به مباحث نظری اصلی پایان‌نامه را ارایه می‌کنیم. عمده مطالب این فصل از پاشا (۱۳۸۰) و جانسون (۲۰۰۹) برگرفته شده‌اند.

۱.۱ مقدمه

در دیدگاه بیزی استنباط آماری، تمامی استنباط‌ها، مبتنی بر توزیع پسین به‌دست می‌آیند. در اغلب موقعیت‌های کاربردی توزیع پسین صورت بسته‌ای ندارد و باید به شیوه‌ای آن را تقریب زد. یک رده از روش‌ها برای تقریب توزیع پسین، روش‌های نمونه‌گیری مونت‌کارلوی زنجیر مارکوف[۱]،(MCMC) به‌ویژه الگوریتم‌های متروپولیس- هستینگز[۲]، (MH) و نمونه‌گیر گیبز[۳]، (GS) می‌باشند. این روش‌ها، بنابر زنجیرهای مارکوف طوری طراحی می‌شوند که با افزایش تعداد تکرارهای نمونه‌گیری یا به عبارتی با به‌روز رسانی حالت زنجیر از جایی به بعد، حالت‌های زنجیر حاصل به عنوان نمونه‌هایی از توزیع مطلوب، یعنی توزیع پسین، در نظر گرفته می‌شوند.

در عمل انتخاب توزیع پیشنهادی[۴] می‌تواند چالش برانگیز باشد، به‌خصوص در مسایلی که بعد فضای حالت بالا باشد. بنابراین به دلیل کاربرد مهم MCMC برای کاربران، و این‌که به‌روز رسانی با بعد بالا دارای محاسبات بیش‌تری است، به‌روز رسانی با بعد پایین‌تر ترجیح داده می‌شود. در روش‌های MCMC، معمول است که به‌جای به‌روز رسانی هم‌زمان همه متغیرها در هر مرحله، فقط یک متغیر (یا بلوکی از متغیرها) به‌روز رسانی شود که به آن به‌روز رسانی مولفه به مولفه[۵] اطلاق می‌شود (نیل و رابرتز، ۲۰۰۶).

در مورد صورت‌های مختلف زنجیرهای متروپولیس–هستینگز زمانی‌که همه متغیرها هم‌زمان به‌روز

---

[۱] Markov Chain Monte Carlo

[۲] Metropolis Hasting

[۳] Gibbs Sampler

[۴] Proposal distribution

[۵] Component-wise updating

می‌شوند پژوهش‌های زیادی انجام شده است. به عنوان چند نمونه می‌توان به منگرسن و توییدی (۱۹۹۶)، کریستنسن و همکاران (۲۰۰۱) و جانسون و گیر (۲۰۱۲) اشاره کرد اما هیچ‌کدام از این نتایج برای زنجیرهای مولفه به مولفه متروپولیس–هستینگز به‌کار گرفته نشده‌اند. تنها روش مولفه به مولفه‌ای که کمی مورد توجه قرار گرفته است، نمونه گیر گیبز است. به عنوان مثال می‌توانید داس و هابرت (۲۰۱۰) و جانسون و جونز (۲۰۱۰) را مشاهده کنید. پس از به‌روز رسانی به روش مولفه به مولفه همه متغیرها (یا بلوکی از متغیرها)، استراتژی‌های مختلفی برای ترکیب آن‌ها از جمله اتصال همگی یا انتخاب یک دنباله تصادفی، به عنوان حالت زنجیر توام همه متغیرها و اسکن تصادفی قابل به‌کارگیری هستند. در ادامه به معرفی استنباط‌های بسامدی و بیزی، زنجیرهای مارکوف ارگودیک و بحث درباره همگرایی آن‌ها و همچنین مدل‌های آمیخته خطی تعمیم‌یافته می‌پردازیم.

اشاره می‌شود که در این فصل برای شروع، برخی مفاهیم اساسی مرتبط با همگرایی مونت‌کارلوی زنجیرهای مارکوف که در فصل‌های بعد پرکاربرد می‌باشند، ارایه شده است.

۲.۱ تعاریف و مفاهیم مقدماتی

تعریف ۱.۲.۱. (فرآیند تصادفی[۶]). فرض کنید $\chi \subset R$ مجموعه ثابتی باشد و T مجموعه‌ای دلخواه. اگر $t \in T$ متغیر شاخصی باشد، که معمولا بیانگر متغیر زمان است، آنگاه دنباله متغیرهای تصادفی $\{X^{(t)} : t \in T\}$ را فرآیند تصادفی با مجموعه اندیس‌گذار T و فضای حالت χ می‌گویند.

تعریف ۲.۲.۱. (زنجیر مارکوف). فرآیند تصادفی $\{X^{(t)} : t \in T\}$ را یک زنجیر مارکوف گویند، اگر

$$P(X^{(k+1)} = a|X^{(k)}, X^{(k-1)}, \ldots, X^{(0)}) = P(X^{(k+1)} = a|X^{(k)}) = K(X^{(k)}, X^{(k+1)}).$$

در آن $K(\cdot,\cdot)$ را هسته انتقال[۷] یا هسته مارکوف می‌نامند. به هسته انتقال مارکوف، چگالی انتقال مارکوف[۸]، (Mtd) نیز گویند.

در زنجیر مارکوف فقط اطلاع از حالت فرآیند در مرحله k، برای تعیین توزیع حالت فرآیند در مرحله $k+1$ کفایت می‌کند و اطلاعات قبل از آن موثر نخواهد بود. از آن‌جا که نمونه بعدی در هر زنجیر فقط به نمونه قبلی آن وابسته است، این عدم وابستگی به گذشته اجازه می‌دهد، تا زنجیرهای مارکوف برای ساده‌سازی مسایل پیچیده مورد استفاده قرار گیرند.

تعریف ۳.۲.۱. (احتمال انتقال). احتمال شرطی $P(X^{(k+1)} = y|X^{(k)} = x)$ را احتمال انتقال یک‌مرحله‌ای (از x در مرحله k ام به y در مرحله‌ی $k+1$ام) می‌نامیم، و در زنجیر مارکوف همگن،(احتمال انتقال یک‌مرحله‌ای که به k بستگی نداشته باشد) احتمال‌های انتقال را با P_{xy} نشان می‌دهیم.

$$P_{xy} = P\{X^{(k+1)} = y|X^{(k)} = x\}, \quad k \geq 0, x, y \in \chi$$

.

[۶] Stochastic process

[۷] Transition kernel

[۸] Mrkov transition density

تعریف ۴.۲.۱. (احتمال انتقال چندمرحله‌ای). فرض کنید $\{X^{(t)} : t \in T\}$ یک زنجیر مارکوف با فضای حالت χ و ماتریس احتمال $P = (p_{xy})$ باشد. به ازای $m, n \geq ۱$، احتمال‌های

$$P\{X^{(n+m)} = y | X^{(m)} = x\} \quad x, y \in \chi$$

را احتمال‌های انتقال n مرحله‌ای می‌گویند و با علامت $p_{xy}^{(n)}$ نشان می‌دهند.

یک زنجیر مارکوف با مشخصات دلخواه ما، از دیدگاه استنباط بیزی، نیاز به داشتن یک سری ویژگی‌هایی است که در ادامه تعریف می‌کنیم.

تعریف ۵.۲.۱. (زنجیر مارکوف مانا). زنجیر مارکوف $\{X^{(t)} : t \in T\}$ با توزیع مانای[۹] $\pi(\cdot)$ است، هرگاه $X^{(t)} \sim \pi(\cdot)$ آنگاه $X^{(t+۱)} \sim \pi(\cdot)$. یعنی حالت زنجیر در هر مرحله، نسخه مشابهی از حالت آغازین زنجیر خواهد بود.

تعریف ۶.۲.۱. فرآیندی ماناست، اگر $\pi_۱ = \pi_۰$، آنگاه به ازای هر $n \geq ۱$، $\pi_n = \pi_۰$.

تعریف ۷.۲.۱. (تحویل‌ناپذیری[۱۰]). اگر فضای حالت زنجیری تنها یک دسته هم‌ارزی داشته باشد، یعنی هر دو حالت دلخواه در دسترس یکدیگر باشند، آن زنجیر را تحویل‌ناپذیر گوییم و یا به عبارت دیگر، K، صرفنظر از مقدار اولیه $X^{(۰)}$، اجازه حرکت بر روی تمام وضعیت‌های فضای حالت زنجیر را خواهد داشت. یا به بیان دیگر، اگر برای همه $x \in \chi$ و $A \in \mathcal{B}$ یک n وجود داشته باشد که $P^n(x, A) > ۰$.

ملاحظه ۸.۲.۱. *(ϕ-تحویل‌ناپذیر). زنجیر مارکوف، برای برخی اندازه[۱۱] ϕ روی $(X, \mathcal{B})$، ϕ-تحویل‌ناپذیر است اگر برای همه $x \in \chi$ و $A \in \mathcal{B}$، $\phi(A) > ۰$ باشد، آنگاه یک n وجود دارد که $P^n(x, A) > ۰$.*

تعریف ۹.۲.۱. (بازگشت‌پذیری[۱۲]). اگر زنجیری بازگشتی باشد با شروع از هر نقطه بالاخره در یک زمان متناهی با احتمال مثبت به آن باز خواهد گشت.

ملاحظه ۱۰.۲.۱. *(شرط تعادل دقیق[۱۳]). زنجیر مارکوف روی فضای حالت χ با توجه به توزیع احتمال π، برگشت‌پذیر است، اگر*

$$\pi(x)P(x, dy) = \pi(y)P(y, dx) \qquad \forall x, y \in \chi.$$

ملاحظه ۱۱.۲.۱. *اگر زنجیر مارکوفی با توجه به $\pi(\cdot)$ برگشت‌پذیر باشد، آنگاه $\pi(\cdot)$ برای زنجیر ماناست (رابرتز و روزنتال، ۲۰۰۴).*

ملاحظه ۱۲.۲.۱. *شرط تعادل دقیق روشی برای حصول اطمینان، از مانایی توزیع π است. اگر این ویژگی برقرار باشد زنجیر مارکوف با توجه به π برگشت‌پذیر است.*

---

[۹] Invariant distribution

[۱۰] Irreduciblity

[۱۱] Measure

[۱۲] Reversiblity

[۱۳] Detailed balance condition

ملاحظه ۱۳.۲.۱. *(برگشت‌پذیر هریس[۱۴]). یک زنجیر مارکوف $\Phi = \{X^{(0)}, X^{(1)}, \dots\}$ برگشت‌پذیر هریس است اگر به ازای هر مقدار اولیه x زنجیر، مجموعه $A \in \mathcal{B}$ که در آن $\mathcal{B}$ یک σ-جبر بورل است با احتمال یک، بی‌نهایت بار ملاقات کند، یعنی $P(X^{(n)} \in A | X^{(0)} = x) \to 1$.*

ملاحظه ۱۴.۲.۱. *(نادوره‌ای[۱۵]). به این معنی است که زنجیر، در جستجوی فضای حالت، گرفتار هیچ چرخشی نشود. به عبارت دیگر زنجیر در یک مجموعه موضعی گیر نکند.*

ملاحظه ۱۵.۲.۱. *(ارگودیک هریس[۱۶]). یک زنجیر مارکف ارگودیک هریس است، اگر ϕ-تحویل‌ناپذیر، نادوره‌ای، هریس برگشت‌پذیر و دارای توزیع مانای $\pi(\cdot)$ برای برخی اندازه ϕ و π باشد.*

تعریف ۱۶.۲.۱. (ارگودیک). فرض کنید زنجیر مارکوف Φ ارگودیک هریس با توزیع مانای π باشد و همچنین فرض کنید به ازای هر تابعی $g : X \to \mathbb{R}$، $E_\pi |g(x)| < \infty$. آن‌گاه به ازای هر مقدار اولیه $x \in \chi$، اگر $n \to \infty$

$$\bar{g}_n = \frac{1}{n} \sum_{i=0}^{n-1} g(X^{(i)}) \to E_\pi g(X) \qquad almost\ surely.$$

ملاحظه ۱۷.۲.۱. *(فاصله تغییرات کل[۱۷]). فرض کنید $\mu(\cdot)$ و $\nu(\cdot)$ دو اندازه روی $(\chi, \mathcal{B})$ باشند. آن‌گاه فاصله تغییرات کل را به‌صورت زیر تعریف می‌کنیم*

$$\|\mu(\cdot) - \nu(\cdot)\| = sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|$$

ملاحظه ۱۸.۲.۱. *فرض کنید زنجیر مارکوف Φ ارگودیک هریس با توزیع مانای $\pi(\cdot)$ باشد. آن‌گاه برای هر مقدار اولیه $x \in \chi$، Φ به $\pi(\cdot)$ در فاصله تغییرات کل همگرا خواهد شد. بنابراین اگر $n \to \infty$*

$$\|P^n(x, \cdot) - \pi(\cdot)\| \to 0.$$

مثال ۱۹.۲.۱. فرض کنید فضای حالت $\chi = \{1, 2, 3\}$ با $\pi\{1\} = \pi\{2\} = \pi\{3\} = \frac{1}{3}$ و فرض کنید $p(2, \{2\}) = p(2, \{1\}) = p(1, \{2\}) = p(1, \{1\}) = \frac{1}{4}$ و $p(3, \{3\}) = 1$ این مثال تحویل‌پذیر است یعنی زنجیر هرگز از حالت ۱ به حالت ۳ نمی‌رود. از طرفی این زنجیر برگشت‌پذیر نیز نمی‌باشد زیرا با احتمال مثبت به تمام نقاط زنجیر حرکت نمی‌کند.

مثال ۲۰.۲.۱. فرآیند $\{X_n : n = 0, 1, 2, \dots\}$ زنجیری مارکوف است، فرض کنید که دارای فضای حالت $\chi = \{0, 1\}$، ماتریس احتمال انتقال $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ و با توزیع $\pi_0(\frac{1}{2}, \frac{1}{2})$ باشد.

اگر ما نشان دهیم که، به ازای هر n، رابطه

$$\pi_1 = \pi_0 P, \pi_2 = \pi_0 P^2 \dots, \pi_n = \pi_0 P^n = \pi_0$$

[۱۴] Harris recurrent

[۱۵] Aperiodic

[۱۶] Harris ergodic

[۱۷] Total variation distance

برقرار باشد، فرآیند ماناست. با محاسبه

$$\pi_۱ = \pi_۰ P = (\frac{۱}{۲}\ \frac{۱}{۲}) \begin{bmatrix} \frac{۱}{۲} & \frac{۱}{۲} \\ \frac{۱}{۲} & \frac{۱}{۲} \end{bmatrix} = (\frac{۱}{۲}\ \frac{۱}{۲})$$

$$\pi_۲ = \pi_۰ P^۲ = (\frac{۱}{۲}\ \frac{۱}{۲}) \begin{bmatrix} \frac{۱}{۲} & \frac{۱}{۲} \\ \frac{۱}{۲} & \frac{۱}{۲} \end{bmatrix} = (\frac{۱}{۲}\ \frac{۱}{۲})$$

اگر این روند محاسبه را تا مرحله n ام ادامه دهیم به این نتیجه می‌رسیم که توزیع $\pi_n = \pi_۰$، درنتیجه مانایی مثال ثابت می‌شود.

۳.۱ دیدگاه‌های بسامدی و بیزی در استنباط آماری

روش‌های آماری را در یک دسته‌بندی کلی می‌توان به دو دسته بسامدی و بیزی تقسیم‌بندی کرد. هرکدام از این چارچوب، منطق و هدف خاص خود را در استنباط آماری دنبال می‌کند (رابرت، ۲۰۰۷). در دیدگاه بسامدی استنباط آماری، پارامتر مورد بررسی ثابت، اما نامعلوم است. در زمینه مدل‌بندی، یکی از شاخه‌های پرطرفدار استنباط‌های بسامدی مبتنی بر درستنمایی است. در مقابل دیدگاه مبتنی بر درستنمایی، دیدگاه بیزی فرض می‌کند که مقدار واقعی پارامتر، تحققی از یک توزیع تصادفی است. به عبارتی پارامتر یک متغیر تصادفی محسوب می‌شود.

مزیت عمده استنباط بیزی این است که پسین همیشه با یک روش واحد به‌دست می‌آید، اطلاعات اولیه درباره پارامتر را از توزیع پیشین[۱۸] و اطلاعات موجود در داده‌های مشاهده شده را توسط تابع درستنمایی ترکیب می‌کند و چگالی پسین[۱۹] را می‌سازد.

فرض کنید $y = (y_۱, \ldots, y_n) \in \mathbb{R}^d$ یک نمونه nتایی، دارای تابع درستنمایی $L(\theta|y)$ باشد به‌طوری‌که $\theta = (\theta_۱, \ldots, \theta_d) \in \Theta \subset \mathbb{R}^d$. همچنین فرض کنید بردار پارامتر θ دارای تابع چگالی پیشین $\pi(\theta)$ باشد. در این‌صورت، بنابر قضیه بیز

$$\pi(\theta|y) = \frac{\pi(\theta) \cdot L(\theta|y)}{\int_\Theta \pi(\theta) \cdot L(\theta|y) d\theta} = c(y) \cdot \pi(\theta) \cdot L(\theta|y)$$

که ترکیب توزیع پیشین و تابع درستنمایی، به تابع توزیع احتمال جدیدی منجر می‌شود که آن‌را توزیع پسین می‌نامیم و تمامی استنباط‌های بیزی، بر اساس آن صورت می‌گیرد. از طرفی در رابطه فوق $c(\mathbf{y})$ ثابت نرمال‌ساز[۲۰] نامیده می‌شود که به پارامتر $\boldsymbol{\theta}$ وابسته نیست و برابر است با

$$c(y) = \frac{۱}{\int_\Theta \pi(\theta) \cdot L(\theta|y) d\theta}$$

شایان ذکر است که تابع چگالی پسین را، زمانی می‌توان محاسبه کرد که تابع درستنمایی قابل محاسبه باشد. توزیع پسین معمولا صورت بسته‌ای ندارد و بسته به انتخاب توزیع پیشین، پیچیده هستند. بنابراین برای استخراج استنباط‌های آماری باید از روش‌های تقریبی، برای تقریب توزیع پسین استفاده کرد.

[۱۸] Perior distribution

[۱۹] Posterior density

[۲۰] Normalizing constant

روش‌های تقریب توزیع پسین، به دو دسته کلی، روش‌های عددی و روش‌های مبتنی بر نمونه‌گیری تقسیم‌بندی می‌شوند. از روش‌های تقریب عددی می‌توان به تقریب لاپلاس[۲۱] (تیرنی و کدین، ۱۹۸۶)، روش‌های مربع‌بندی[۲۲] (پینیرو و بیتس، ۱۹۹۵) و تقریب نیوتن-رافسون اشاره کرد. از معروف‌ترین روش‌های مبتنی بر نمونه گیری که به روش‌های شبیه‌سازی مونت‌کارلویی نیز معروف هستند، می‌توان روش‌های نمونه‌گیری نقاط مهم[۲۳] (هستربرگ، ۱۹۸۷)، رد و پذیرش[۲۴] (کسلا و برگر،۲۰۰۲) و مونت‌کارلوی زنجیر مارکوفی (هستینگز، ۱۹۷۰؛ رابرت و کسلا، ۲۰۰۴) نام برد.

هر یک از روش‌های ذکر شده دارای معایب قابل توجهی هستند. مشکل اساسی روش‌های انتگرال‌گیری عددی در شناسایی نواحی مهم (نواحی چگال‌تر) برای تابعی که باید انتگرال‌گیری شود، می‌باشد. انتگرال‌گیری‌های عددی برای نواحی با بعد حداکثر ۶ خوب عمل می‌کنند و در غیر این‌صورت با خطای زیادی مواجه می‌شوند، زیرا با افزایش بعد انتگرال، با تجمع خطاها مواجه می‌شویم. افزایش بعد انتگرال، افزایش محاسبات را نیز به همراه خواهد داشت (بریسلو و کلیتون، ۱۹۹۳).

۴.۱ روش‌های نمونه‌گیری مونت‌کارلوی زنجیر مارکوف

روش‌های مبتنی بر نمونه‌گیری برای تقریب توزیع پسین انتخاب معمول هستند. در روش های مبتنی بر نمونه گیری، قانون قوی اعداد بزرگ و قضیه ارگودیک[۲۵] (بیرخوف، ۱۹۴۲)، همگرایی برآوردگرهای حاصل از نمونه تولیدشده، به کمیت‌های مورد نظر از توزیع پسین را تضمین می‌کند. در ساخت روش تقریبی باید ابتدا الگوریتم‌های نمونه‌گیری MCMC را توضیح دهیم.

الگوریتم‌های مونت‌کارلوی زنجیر مارکوف

در مواردی که صورت بسته توزیع پسین در دسترس نیست و بعد پارامترهایش بالا باشد، برای محاسبه تقریبی آن، از روش‌های تقریبی استفاده می‌شود. یکی از روش‌های تقریب توزیع پسین استفاده از روش‌های شبیه‌سازی مونت‌کارلویی می‌باشد. برای رفع این مشکل الگوریتم‌های MCMC (متروپولیس و همکاران، ۱۹۵۳؛ هستینگز، ۱۹۷۰) معرفی شدند که به‌طور غیر مستقیم به تولید نمونه از یک توزیع احتمالی مفروض می‌پردازند.

مکانیسم الگوریتم‌های MCMC، مبتنی بر تولید نمونه وابسته از یک زنجیر مارکوف است که توزیع مانای آن، همان توزیع پسین موردنظر می‌باشد. در این صورت، اگر حجم نمونه مونت‌کارلویی تولید شده، به اندازه کافی بزرگ باشد، از جایی به بعد، حالت‌های زنجیر حاصل به عنوان نمونه‌هایی از توزیع مطلوب، یعنی توزیع پسین، در نظر گرفته می‌شود.

---

[۲۱] Laplace approximation

[۲۲] Quadrature methods

[۲۳] Importance sampling

[۲۴] Reject-accept sampling

[۲۵] Ergodic theorem

برای معرفی جزییات الگوریتم‌های MCMC، ابتدا مساله انتگرال‌گیری مونت‌کارلویی را تشریح می‌کنیم.

انتگرال‌گیری مونت‌کارلویی

فرض کنید بخواهیم انتگرال

$$I = \int h(x)dx = \int g(x) \cdot f(x)dx$$

را محاسبه کنیم، که در آن $g(\cdot)$، تابعی از متغیر تصادفی X با تابع چگالی $f(\cdot)$ است. چنان‌چه نتوان این انتگرال را با محاسبات معمولی حل کرد، می‌توان از روش‌های شبیه‌سازی برای تقریب آن استفاده نمود.

برای این منظور ابتدا مساله انتگرال گیری را به یک مساله محاسبه امید ریاضی بر حسب یک تابع چگالی، مانند $f(\cdot)$ تبدیل می‌کنیم. یعنی

$$I = \int h(x)dx = \int g(x) \cdot f(x)dx = E_f[g(x)]$$

آن‌گاه برآورد مونت کارلوی I، برابر میانگین یک نمونه به‌دست آمده از توزیع f، وقتی که حجم نمونه بزرگ باشد، خواهد بود. بنابراین مساله انتگرال‌گیری به مساله یافتن راهی برای تولید نمونه از چگالی هدف $f(\cdot)$، تبدیل می‌شود.

در واقع برای محاسبه انتگرال I لازم نیست از یک نمونه (مستقل) مستقیما تولید شده از توزیع $f(\cdot)$، استفاده شود. می‌توان، بدون تولید مستقیم از $f(\cdot)$، یک نمونه (وابسته)، $X_1, \ldots, X_n$ را با استفاده از یک زنجیر مارکوف ارگودیک با توزیع مانای $f(\cdot)$، به‌دست آورد.

تولید نمونه به روش غیرمستقیم

یک راه‌حل تولید غیر مستقیم نمونه‌ها، استفاده از روش‌های MCMC است. به‌وسیله زنجیرهای مارکوف، به‌جای شبیه‌سازی نمونه‌های مستقل از توزیع پسین، نمونه‌ای وابسته با توزیع مانا شبیه‌سازی می‌شود. این روش، ترکیبی از نمونه‌گیری مونت‌کارلویی و زنجیرهای مارکوف است که به روش نمونه‌گیری MCMC مشهور شد. با این‌که کشف MCMC به مطالعات متروپولیس و همکاران (۱۹۵۳) و هستینگز (۱۹۷۰) برمی‌گردد، اما استفاده شایع آن به عنوان ابزاری برای اجرای استنباط‌های بیزی با کارگلفاند و اسمیت (۱۹۹۰) شروع شد. برای مشاهده تاریخچه پیدایش و گسترش الگوریتم‌های MCMC به رابرت و کسلا (۲۰۱۱) مراجعه کنید.

برای توصیف الگوریتم‌های MCMC، دنباله‌ای از متغیرهای تصادفی وابسته

$$X^{(0)}, X^{(1)}, \ldots, X^{(k)}, \ldots$$

که هر یک روی فضای حالت χ تعریف شده‌اند، در نظر بگیرید. هدف اصلی الگوریتم MCMC ساخت زنجیر مارکوفی است که توزیع مانای آن توزیع هدف π باشد. اطمینان از وجود توزیع مانا، نیازمند آن است که زنجیر مارکوف، دارای چندین ویژگی باشد که، در بخش تعاریف این فصل، به آن‌ها اشاره کردیم.

بنابراین تحت این ویژگی‌ها، توزیع مانای π یک توزیع حدی هم خواهد بود. به این معنی که $X^{(k)}$ در توزیع به π همگرا می‌شود. یعنی اگر $k \to \infty$

$$X^{(k)} \xrightarrow{\mathrm{d}} \pi$$

لازم به ذکر است که نقطه شروع تاثیری بر این همگرایی ندارد. نتیجه این ویژگی همگرایی، قضیه ارگودیک است که بیان می‌کند، میانگین تجربی تابع دلخواه $g(x)$ حاصل از نمونه‌های وابسته تولیدی از زنجیر مارکوف، با احتمال یک، به مقدار واقعی موردنظر همگرا می‌شود. یعنی

$$\frac{1}{n}\sum_{k=1}^{n} g(X^{(k)}) \xrightarrow{\text{a.s.}} E[g(X)]$$

که این قضیه مشابه قانون قوی اعداد بزرگ برای نمونه‌های مستقل و هم‌توزیع است.

اکنون سوالی که مطرح می‌شود، چگونگی ساخت چنین زنجیری با توزیع مانای مورد نظر ما است، که روش‌های MCMC برای پاسخ به این سوال ارایه شده‌اند.

به‌طور کلی الگوریتم‌های زیادی برای تولید زنجیرهای مارکف با یک توزیع مانای خاص وجود دارد. از میان این الگوریتم‌ها دو الگوریتم نمونه‌گیر گیبز (گمان و گمان، ۱۹۸۴ ؛ گلفاند و اسمیت، ۱۹۹۰) و متروپولیس-هستینگز (متروپولیس و همکاران، ۱۹۵۳؛ هستینگز، ۱۹۷۰) در ادبیات آماری مقبول‌تر هستند. الگوریتم گیبز بر توزیع‌های پسین شرطی پارامترها بنا شده است و می‌بایست برای استفاده از آن، توزیع‌های شرطی کامل پارامترها، دست کم به صورت متناسب مشخص شده باشند. در غیر این‌صورت می‌توان از الگوریتم متروپولیس-هستینگز کمک گرفت.

از طرفی الگوریتم متروپولیس-هستینگز دارای مشکلات متعددی نظیر تشخیص همگرایی زنجیر و زمان طولانی محاسبات می‌باشد. اما نمونه‌گیر گیبز این‌گونه مسایل را به دنباله‌ای از مسایل ساده‌تر تقسیم می‌کند و اجرای آن ساده‌تر است. البته ممکن است این دنباله‌ها زمان زیادی برای همگرایی نیاز داشته باشند ولی با این وجود، این الگوریتم مفید وجالب توجه است، زیرا این اجازه را می‌دهد که یک نمونه از توزیع پسین توام تولید کنیم. در ادامه دو الگوریتم متروپولیس-هستینگز و نمونه‌گیر گیبز را به‌طور مختصر شرح می‌دهیم.

۵.۱ الگوریتم متروپولیس-هستینگز

الگوریتم متروپولیس-هستینگز ابتدا توسط متروپولیس و همکاران (۱۹۵۳) پیشنهاد شد و سپس توسط هستینگز (۱۹۷۰) به‌طور کامل تدوین یافت. این الگوریتم ساختاری مشابه با ساختار رد و پذیرش دارد و مشاهداتی را که توسط یک توزیع پیشنهادی به عنوان نمونه‌های ممکن از توزیع پسین پیشنهاد می‌شوند، بر اساس یک قاعده احتمالاتی مورد ارزیابی قرار داده، و با احتمال معینی می‌پذیرد. بر این اساس برای استفاده از این الگوریتم، به یک توزیع پیشنهادی نیاز است که انتخاب آن بر دقت و سرعت همگرایی الگوریتم موثر است. در این الگوریتم فضای حالت زنجیر $\chi \in \mathbb{R}^q$ که $q \geq 1$ و ϖ یک اندازه احتمال باشد، را در نظر بگیرید و فرض کنید $X^{(k)} = x$ نشان‌دهنده حالت جاری زنجیر، و ϖ دارای

تابع مانای $\pi(\cdot)$ باشد و همچنین $p(\cdot,\cdot)$ نشان‌دهنده چگالی پیشنهادی باشد. الگوریتم MH برای به‌روز رسانی $X^{(k+1)}$ به‌صورت زیر به‌دست می‌آید:

۱) تولید y از تابع چگالی پیشنهادی $p(y|x)$

۲) محاسبه احتمال پذیرش $\alpha(x,y)$ که در آن

$$\alpha(x,y) = min\left\{1, \frac{\pi(y)p(x|y)}{\pi(x)p(y|x)}\right\}$$

۳) قرار دادن $X^{(k+1)} = y$ با احتمال $\alpha(x,y)$ و قرار دادن $X^{(k+1)} = x$ با احتمال $1 - \alpha(x,y)$.

از ویژگی‌های مطلوب این الگوریتم آن است که شناخت توزیع پسین به صورت متناسب کفایت می‌کند. این الگوریتم، با شرط داشتن توزیع پسین، مجموعه‌ای از متغیرهای تصادفی $\{X^{(k+1)}\}$ ، برای $k = 1, 2, \ldots$، به‌طور پی‌درپی تولید می‌کند. برای تولید این نمونه‌ها، این الگوریتم، ما را ملزم به تعیین چگالی پیشنهادی $p(y|x)$ می‌کند. این تابع، تابع چگالی x، احتمال در زمان $k+1$، با توجه به مقدار آن در زمان k است.

نسبت موجود در رابطه فوق هستینگز[۲۶]نامیده می‌شود. اگر توزیع پیشنهادی مقادیری نزدیک به x را پیشنهاد کند، تقریبا همیشه پیشنهادها پذیرفته می‌شوند اما زنجیر خیلی کند حرکت می‌کند و زمان زیادی طول می‌کشد تا الگوریتم همگرا شود. همچنین اگر توزیع پیشنهادی مقادیر دور از x را پیشنهاد کند، پیشنهادها تقریبا همیشه رد می‌شوند و باز هم زنجیر به کندی همگرا می‌شود. با کمی تنظیم می‌توان توزیع پیشنهادی را به نحوی تعدیل کرد که پیشنهادها مکررا رد یا قبول نشوند و الگوریتم از نرخ پذیرش مناسبی برخوردار باشد (رابرتز و همکاران، ۱۹۹۷).

دو حالت خاص الگوریتم MH مستقل[۲۷] (IMH) و الگوریتم MH قدم‌زدن تصادفی [۲۸] (RWMH) را در ادامه معرفی می‌کنیم.

۱.۵.۱ الگوریتم متروپولیس-هستینگز مستقل

در حالتی که مقدار چگالی پیشنهادی جدید، از مقدار جاری زنجیر، یعنی $X^{(k)}$، مستقل باشد، حالت خاصی از الگوریتم اتفاق می‌افتد که الگوریتم حاصل با این انتخاب را MH مستقل گویند. در این حالت $p(y|x) = p(y)$. روند الگوریتم به صورت زیر است:

۱) تولید y از تابع چگالی پیشنهادی $p(y)$

۲) محاسبه احتمال پذیرش $\alpha(x,y)$ که در آن

$$\alpha(x,y) = min\left\{1, \frac{\pi(y)p(x)}{\pi(x)p(y)}\right\}$$

[۲۶] Hastings

[۲۷] Independent Metropolis-Hastings algorithm

[۲۸] Random Walk Metropolis-Hastigs algorithm

۳) قرار دادن $X^{(k+۱)} = y$ با احتمال $\alpha(x, y)$ و قرار دادن $X^{(k+۱)} = x$ با احتمال $۱ - \alpha(x, y)$.

بنابراین، زمانی‌که چگالی پیشنهادی $p(\cdot)$ وابسته به حالت جاری نباشد زنجیر یک نمونه متروپولیس-هستینگز مستقل (MHIS) است. برای اطلاع از ویژگی‌های این نوع الگوریتم می‌توانید به رابرت و کسلا (۲۰۱۰) مراجعه کنید.

۲.۵.۱ الگوریتم متروپولیس -هستینگز قدم زدن تصادفی

در این الگوریتم مقدار جدید پیشنهادی بر طبق مقدار شبیه‌سازی شده قبلی تولید می‌شود، که در آن یک کاوش در همسایگی[۲۹] مقدار جاری زنجیر صورت می‌گیرد.

اجرای این ایده با شبیه‌سازی y طبق $y = x + \epsilon_k$ است که در آن ϵ_k یک خطای تصادفی مستقل از x با تابع چگالی $p(\cdot)$ است. در این حالت توزیع پیشنهادی عبارت است از:

$$p(y|x) = p(x - y) = p(y - x)$$

در صورتی‌که توزیع $p(\cdot)$ حول صفر متقارن باشد (مثلا در توزیع نرمال یا نرمال چندمتغیره، با میانگین صفر و یا توزیع یکنواخت استاندارد) زنجیر مارکوف متناظر با آن، یک فرآیند قدم‌زدن تصادفی است. که الگوریتم آن به صورت زیر است:

۱) تولید y از تابع چگالی پیشنهادی $p(y - x)$

۲) محاسبه احتمال پذیرش $\alpha(x, y)$ که در آن

$$\alpha(x, y) = min\left\{۱, \frac{\pi(y)}{\pi(x)}\right\}$$

۳) قرار دادن $X^{(k+۱)} = y$ با احتمال $\alpha(x, y)$ و قرار دادن $X^{(k+۱)} = x$ با احتمال $۱ - \alpha(x, y)$.

متذکر می‌شویم که احتمال پذیرش به $p(\cdot)$ وابسته نیست. این بدین معنی است که، برای جفت (x, y) احتمال پذیرش y در حالتی که از نرمال یا کوشی تولید شود، مشابه است. اما باید توجه داشت که تغییر $p(\cdot)$ منجر به دامنه‌های مختلف مقادیر، برای y و در نتیجه نرخ‌های پذیرش متفاوت خواهند شد. بنابراین، انتخاب p بر عملکرد الگوریتم موثر است.

۶.۱ نمونه‌گیر گیبز

روش نمونه‌گیری گیبز ابتدا توسط گمان و گمان (۱۹۸۴) در پردازش مدل‌های تصویری معرفی شد. سپس توسط گلفاند و اسمیت (۱۹۹۰) به عنوان روشی بر مبنای نمونه‌گیری برای محاسبه چگالی‌های حاشیه‌ای در چارچوب استنباط بیزی پیشنهاد شد. توضیح واضح و ساده این الگوریتم به‌وسیله کسلا و جورج (۱۹۹۲) بیان شده است.

[۲۹] Local exploration

در رهیافت بیزی هنگامی‌که امکان انجام استنباط‌های تحلیلی پسینی به دلیل ناشناخته یا پیچیده بودن توزیع پسین وجود ندارد، از روش‌های مونت‌کارلویی زنجیر مارکوف که مبتنی بر نمونه‌گیری از توزیع پسین هستند، برای تقریب کمیت‌های مورد علاقه پسینی استفاده می‌شود. الگوریتم گیبز از توزیع‌های پسین شرطی پارامترهای مختلف، به شرط سایر پارامترها که اصطلاحا توزیع پسین شرطی کامل[۳۰] نامیده می‌شوند، برای نمونه‌گیری از توزیع پسین هدف استفاده می‌کند.

۱.۶.۱ نمونه‌گیر گیبز دومرحله‌ای

نمونه‌گیر گیبز دومرحله‌ای، یک زنجیر مارکوف را از توزیع توام به صورت زیر ایجاد می‌کند. اگر فرض کنیم متغیرهای تصادفی X و Y دارای تابع چگالی توام $p(x,y)$ با چگالی‌های شرطی $p(x|y)$ و $p(y|x)$ باشند، نمونه‌گیر گیبز دومرحله‌ای $(X^{(k)},Y^{(k)})$ را به ازای $k=۱,۲,\ldots$ را به صورت زیر تولید می‌کند.

با فرض $X^{(۰)}=x$ برای $k=۱,۲,\ldots$ تولید می‌کنیم

۱) $X^{(k+۱)}\sim P(x|y^{(k)})$

۲) $Y^{(k+۱)}\sim P(y|x^{(k+۱)})$

در صورتی‌که شبیه‌سازی در هر دو مرحله ساده و تولید نمونه از آن‌ها ساده باشد، اجرای الگوریتم آسان است یا به عبارت دیگر، در صورتی‌که چگالی‌های شرطی صورت بسته‌ای داشته باشند، شبیه‌سازی به سادگی انجام می‌شود.

بنابراین در صورتی‌که نمونه‌گیری مستقیم قابل انجام نباشد، تقریب‌های مونت‌کارلو می‌تواند مورد استفاده قرار گیرد. به راحتی در می‌یابیم که اگر $(X^{(k)},Y^{(k)})$ از توزیع $p(x,y)$ باشد، در این‌صورت $(X^{(k+۱)},Y^{(k+۱)})$ نیز از توزیع $p(x,y)$ خواهد بود. به‌عبارت دیگر توزیع مانای این زنجیر، $p(x,y)$ است.

۲.۶.۱ نمونه‌گیر گیبز چندمرحله‌ای

نمونه‌گیر گیبز چندمرحله‌ای تعمیمی از نمونه‌گیر گیبز دومرحله‌ای است. فرض کنید که برای $m>۱$ ، متغیر تصادفی $\mathbf{Y}$ به صورت $\mathbf{Y}=(Y_۱,\ldots,Y_m)$ نوشته شود. مولفه‌های Y_i ممکن است تک بعدی یا چندبعدی باشند. می‌توان برای $i=۱,۲,\ldots,m$ به صورت زیر شبیه‌سازی کرد:

$$Y_i|y_۱,y_۲,\ldots,y_{i-۱},y_{i+۱},\ldots,y_m\sim p_i(y_i|y_۱,y_۲,\ldots,y_{i-۱},y_{i+۱},\ldots,y_m)$$

الگوریتم نمونه‌گیر گیبز تحت انتقال از $Y^{(k)}$ به $Y^{(k+۱)}$ با تکرار $k=۱,۲,\ldots$ و با فرض کردن $\mathbf{y}^{(k)}=(y_۱^{(k)},y_۲^{(k)},\ldots,y_m^{(k)})$ طی گام‌های زیر تولید می‌شود:

۱) $Y_۱^{(k+۱)}\sim p_۱(y_۱|y_۲^{(k)},\ldots,y_m^{(k)})$

[۳۰] Full conditional

۲) $Y_2^{(k+1)} \sim p_2(y_2|y_1^{(k+1)}, y_3^{(k)}, \ldots, y_m^{(k)})$

$\vdots$

m) $Y_m^{(k+1)} \sim p_m(y_m|y_1^{(k+1)}, \ldots, y_{m-1}^{(k+1)})$

چگالی‌های $p_1, \ldots, p_m$ چگالی‌های شرطی کامل نامیده می‌شوند. یک خصوصیت نمونه‌گیر گیبز این است که تنها چگالی‌ها برای شبیه‌سازی مورد استفاده قرار می‌گیرند. بنابراین، حتی مساله‌ای با ابعاد بالا را می‌توان به چند مساله با بعدهای کوچک‌تر تقسیم کرد که این مزیت روش نمونه‌گیری گیبز به حساب می‌آید.

از طرفی هر اندازه میزان خودهمبستگی بین نمونه‌های تولید شده زیاد باشد، الگوریتم برای همگرایی، نیازمند تعداد تکرارهای بیش‌تری است. اما در حالت کلی همگرایی الگوریتم‌های مونت‌کارلوی زنجیر مارکوف به انتخاب توزیع‌های پیشین پارامترها بستگی دارد.

۷.۱ الگوریتم متروپولیس- هستینگز درون گیبز

نمونه‌گیر گیبز را در صورتی‌که برخی از چگالی‌های شرطی کامل را نتوان به سادگی شبیه‌سازی کرد، به آسانی با تنظیماتی می‌توان تعمیم داد. اگر تحت یک مجموعه از چگالی‌های شرطی کامل $p_1, \ldots, p_m$ برخی از p_iها غیر متداول باشند، نتایج نمونه‌گیر گیبز در صورتی‌که راهکار متروپولیس تحت گیبز را به‌کار بگیریم، به مخاطره نمی‌افتد. مثلا فرض کنید شبیه‌سازی از

$$Y_i^{(k+1)} \sim p_i(y_i|y_1^{(k+1)}, \ldots, y_{i-1}^{(k+1)}, y_{i+1}^{(k)}, \ldots, y_m^{(k)})$$

مشکل باشد. در این حالت می‌توان برای تولید نمونه از آن (در یک مرحله از مراحل نمونه‌گیر گیبز) از الگوریتم متروپولیس-هستینگز استفاده کرد. در این حالت این گام با یک الگوریتم متروپولیس-هستینگز که توزیع مانا آن $p(y_i|y_1^{(k+1)}, \ldots, y_{i-1}^{(k+1)}, y_{i+1}^{(k)}, \ldots, y_m^{(k)})$ است، جایگزین می‌شود. این نوع الگوریتم ترکیبی از دو الگوریتم گیبز و متروپولیس-هستینگز است که به آن الگوریتم متروپولیس-هستینگز درون گیبز[۳۱] گویند. در حالتی که شبیه‌سازی از یک چگالی شرطی کامل بسیار مشکل باشد، می‌توان به همین ترتیب مراحل مختلفی از نمونه‌گیر گیبز را با الگوریتم‌های متروپولیس-هستینگز ترکیب کرد.

۸.۱ زنجیر مارکوف ارگودیک و همگرایی آن

معمولی‌ترین کاربرد MCMC برآورد $E_\pi g(X) = \int_\chi g(x)\pi(dx)$ است که g یک تابع حقیقی مقدار و π تابع انتگرال‌پذیر روی χ است. زیرنویس π میانگین امید ریاضی است که $X^{(1)} \sim \pi$ فرض شده است. روش اساسی MCMC مستلزم به ساخت یک زنجیر مارکوف Φ روی χ با داشتن π به عنوان توزیع مانا است. آن‌گاه ما x را برای یک تعداد متناهی از مراحل، که n تا است شبیه‌سازی می‌کنیم و مقادیر مشاهده شده را برای برآورد $E_\pi g$ با یک میانگین نمونه‌ای، $\bar{g}_n = \frac{1}{n}\sum_{i=0}^{n-1} g(\theta_i)$، به کار می‌بریم.

[۳۱] Metropolis-within-Gibbs algorithm

تحت ارگودیک هریس برآورد مونت‌کارلو $\bar{g}_n$ به $E_\pi g$ با احتمال یک زمانی‌که $n \to \infty$ میل می‌کند، همگرا می‌شود. این شرایط تضمین می‌کند که زنجیر مارکوف به توزیع هدف تحت فاصله تغییرات کل همگرا می‌شود. یعنی همگرایی‌اش، همگرایی فاصله تغییرات کل است.

یک زنجیر مارکوف با یک نرخ هندسی به توزیع مانای خود همگرا شود را ارگودیک هندسی گویند. که در ادامه به‌طور دقیق تعریف می‌شود. ارگودیک هندسی به سه دلیل حایز اهمیت است: اول، همگرایی زنجیر مارکوف را تضمین می‌کند. این ضمانت از آن جهت مهم است که به نتایج شبیه‌سازی موثر در کوتاه مدت دسترسی پیدا می‌کنیم؛ دوم، ارگودیک هندسی یک شرط کلیدی برای وجود قضیه حد مرکزی است (جونز، ۲۰۰۴)؛ سوم، برای برآورد خطای استاندارد مونت‌کارلویی قابل اطمینان مورد نیاز است.

از نقطه نظر عملی، ارگودیک هندسی می‌تواند یک برآورد صحیح از پارامتر $E_\pi g$ در یک شبیه‌سازی به اندازه کافی طولانی و یک خطای مونت‌کارلوی نامعلوم، $\bar{g}_n - E_\pi g$، نتیجه دهد، در حالی‌که ارزیابی این خطا به‌طور مستقیم غیرممکن است.

تعریف ۱.۸.۱. (ارگودیک یکنواخت[۳۲]). اگر یک تابع حقیقی مقدار $M(x)$ روی X و $0 < t < 1$ وجود داشته باشد به‌طوری‌که

$$\|P^n(x,\cdot) - \pi(\cdot)\| \leq M(x)t^n$$

اگر $M(x)$ متناهی باشد آن‌گاه Φ ارگودیک یکنواخت است.

تعریف ۲.۸.۱. یک زنجیر مارکوف با توزیع مانای $\pi(\cdot)$ ارگودیک یکنواخت است اگر و تنها اگر برای برخی $n \in N$، $sup_{x\in\chi}\|P^n(x,\cdot) - \pi(\cdot)\| < \frac{1}{2}$ ، (رابرتز و روزنتال، ۲۰۰۴).

مثال ۳.۸.۱. اگر فضای حالت به‌صورت $\chi = \{1,2\}$ با $P(1,\{1\}) = P(2,\{2\}) = 1$ باشد و $\pi(\cdot)$ روی χ یکنواخت تعریف شود، آن‌گاه

$$\|P^n(x,\cdot) - \pi(\cdot)\| = E|P^n(x,\cdot) - \pi(\cdot)| = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$$

با این محاسبه به این نتیجه می‌رسیم که شرط ارگودیک یکنواخت برقرار نمی‌باشد.

تعریف ۴.۸.۱. (ارگودیک هندسی[۳۳]). اگر یک تابع حقیقی مقدار $M(x)$ روی χ و $0 < t < 1$ وجود داشته باشد به‌طوری‌که

$$\|P^n(x,\cdot) - \pi(\cdot)\| \leq M(x)t^n$$

اگر $M(x)$ متناهی نباشد آن‌گاه Φ ارگودیک هندسی است.

مثال ۵.۸.۱. فرض کنید $log\pi(x,y) = -(x^2 + x^2y^2 + y^2)$ که $x, y \in \mathbb{R}^2$. منگرسن و توییدی (۱۹۹۶)، یارنر و هنسن (۲۰۰۰) و رابرتز و توییدی (۱۹۹۶) نشان دادند که یک نمونه متروپولیس-هستینگز قدم‌زدن تصادفی با توزیع هدف π نمی‌تواند ارگودیک هندسی باشد، زیرا ویژگی‌های تعیین شده ارگودیک هندسی در متروپولیس-هستینگز قدم‌زدن تصادفی برقرار نمی‌باشد. لازم به ذکر است زمانی‌که

[۳۲] Uniform ergodicity

[۳۳] Geometric ergodicity

شیوه به‌روز رسانی به روش مولفه به مولفه در نظر گرفته شده است ارگودیک هندسی متروپولیس-هستینگز قدم‌زدن تصادفی ایجاد شده است، که ما در فصل سوم این پایان‌نامه به آن می‌پردازیم.

یک شرط ضعیف‌تر از ارگودیک یکنواخت، ارگودیک هندسی است. برای مطالعه تاریخچه آن، می‌توانید به ناملین، (۱۹۸۴) و مین و توییدی (۱۹۹۳) مراجعه کنید. اختلاف ارگودیک هندسی و یکنواخت این است که، مقدار ثابت M ممکن است به مقدار اولیه x وابسته باشد. البته اگر فضای حالت χ متناهی باشد، آن‌گاه همه زنجیرهای تحویل‌ناپذیر و نادوره‌ای ارگودیک هندسی (در حقیقت ارگودیک یکنواخت) هستند (منگرسن و توییدی، ۱۹۹۶).

ارگودیک یکنواخت، ارگودیک هندسی را نتیجه می‌دهد. اگر زنجیر ارگودیک یکنواخت باشد، این اطمینان حاصل می‌شود که زنجیر کاملا به توزیع مانای خود همگرا است و قضیه حد مرکزی[۳۴] برقرار می‌باشد. ارگودیک یکنواخت یک شرط قوی است که، معمولا در عمل خیلی کم اتفاق می‌افتد. جانشینی برای ارگودیک یکنواخت، ارگودیک هندسی است که، زنجیر مارکوف به توزیع مانای خود با یک نرخ هندسی همگرا می‌شود. ارگودیک هندسی در واقع یک نسخه ضعیف‌تر از ارگودیک یکنواخت است ولی در اکثر موارد می‌توان فرآیندهای تصادفی را که ارگودیک هندسی هستند، مورد بررسی قرار داد.

ما خطای مونت کارلو را نمی‌دانیم اما می‌توانیم از توزیع نمونه‌ای تقریبی، آن را به‌دست آوریم و خطای مونت‌کارلو را به‌وسیله برآورد واریانس توزیع مجانبی $\bar{g}_n$، محاسبه کنیم. اگر قضیه حد مرکزی برای زنجیرهای مارکوف برقرار باشد و اگر $n \to \infty$ آن‌گاه

$$\sqrt{n}(\bar{g}_n - E_\pi g) \xrightarrow{\mathrm{d}} N(0, \sigma_g^2) \qquad (۱.۱)$$

که در آن $\sigma_g^2 \in (0, \infty)$.

با توجه به CLT ما می‌توانیم خطای مونت‌کارلو $\bar{g}_n$ را، به‌وسیله برآورد واریانس σ_g^2 بررسی کنیم. برآورد σ_g^2 را با $\hat{\sigma}_g^2$ نشان می‌دهند. تکنیک‌های بسیاری برای برآورد σ_g^2 معرفی شده است. برای مثال، روش‌های طیفی[۳۵] و میانگین دسته‌ایی[۳۶] واریانس برآوردگرها، در خروجی MCMC می‌توانند مناسب باشند (اتچد، ۲۰۱۱؛ فلگال و جونز، ۲۰۱۰ و جونز و همکاران، ۲۰۰۶). در بین این سه روش، میانگین دسته‌ای برای پیاده‌سازی ساده‌تر می‌باشد.

معیارهای بررسی آمیختگی الگوریتم‌های MCMC

در کنار استفاده شایع از روش‌های نمونه‌گیری MCMC در استنباط‌های بیزی، دو چالش مهم تشخیص همگرایی و ویژگی‌های آمیختگی زنجیر و زمان محاسبات مطرح هستند. برای شناسایی همگرایی زنجیرهای MCMC، معیارهای مختلفی پیشنهاد شده‌اند.

اگر $\delta > 0$، $E_\varpi |g(X)|^{2+\delta} < \infty$ و زنجیر مارکوف ارگودیک هندسی باشد آن‌گاه قضیه حد مرکزی رابطه (۱.۱)، برقرار است، در این حالت می‌توانیم با استفاده از شبیه‌سازی احیاکننده برآورد $E_\pi g$ و خطای استاندارد مجانبی قابل اطمینان مونت‌کارلو را به‌دست آوریم.

[۳۴] Central Limited Theorem

[۳۵] Spectral methods

[۳۶] Batch means

$t_*\sigma_g/\sqrt{n}$، نصف پهنای فاصله اطمینان مجانبی برای $E_\varpi g$ است که t_* بیان کننده یک کمیت مناسب و n نشان‌دهنده طول شبیه‌سازی مونت‌کارلوی زنجیر مارکوف می‌باشد. پهنای فاصله اطمینان، می‌تواند برای تعیین تعداد تکرارهای مورد نیاز برای رسیدن به سطح مطلوبی از دقت قابل استفاده باشد (فلگال، هاران و جونز، ۲۰۰۸؛ فلگال و جونز، ۲۰۱۱ و جونز و همکاران، ۲۰۰۶).

کارایی نسبی زنجیر مارکوف برای بهره‌وری از یک نمونه تصادفی ϖ، به‌وسیله معیار همبستگی یکپارچه زمان[۳۷]، ACT ، قابل اندازه‌گیری است که به‌صورت

$$ACT = \frac{\sigma_g^2}{Var_\varpi(g(X))}.$$

تغییرات برآورد مونت‌کارلو از برآوردی بر اساس یک نمونه تصادفی با همان اندازه مقایسه می‌شود. در عمل $Var_\varpi(g(X))$ و σ_g^2 نامعلوم هستند. با این حال برآورد سازگار $Var_\varpi(g(X))$ به‌وسیله واریانس نمونه‌ای $\widehat{Var}_\varpi(g(X))$ به‌دست می‌آید و از آن‌جایی که زنجیر ارگودیک هندسی است، برآوردگر میانگین دسته‌ایی سازگار، که با $\hat{\sigma_g}^2$ نشان داده می‌شود، یک برآوردگر سازگار از σ_g^2 ایجاد می‌کند. معیار ACT تعداد تکرارهای مورد نیاز نمونه‌گیر گیبز، برای هر نمونه تصادفی انتخاب شده، به منظور رسیدن به همان سطح از دقت در برآورد $E_\varpi g(X)$ را نشان می‌دهد.

قابل ذکر است که روش میانگین دسته‌ای برای ساخت خطای استاندارد مونت‌کارلو سازگار، به‌کار می‌رود. در این روش فرض می‌کنیم a تعداد دسته‌ها، هر کدام به اندازه b، برای یک زنجیر مارکوف به‌طول n باشد. آن‌گاه $n = ab$. برای $k = 1, 2, \ldots, a$ فرض کنید

$$S_k = \frac{1}{b} \sum_{i=(k-1)b}^{kb-1} g(X^{(i)})$$

میانگین تابع زنجیر در k دسته باشد. برآورد میانگین دسته‌ایی σ_g^2

$$\hat{\sigma}_g^2 = \frac{b}{a-1} \sum_{k=1}^{a} (S_k - \bar{g}_n)^2$$

حاصل می‌شود(جونز و همکاران، ۲۰۰۶).

با توجه به اندازه نمونه، کیفیت برآوردهای مونت کارلو به‌وسیله میانگین مربع خطا[۳۸] (MSE) می‌تواند ارزیابی شود. برای برآورد (MSE) ما m تکرار مستقل از هر زنجیر، هر کدام با طول n، اجرا می‌کنیم و برآوردهای مستقل $\bar{g}_n^{(1)}, \ldots, \bar{g}_n^{(m)}$ و یک برآورد مستقل براساس یک اجرای طولانی از یک زنجیری که از قبل مشخص شده است، که به آن $\bar{g}^*$ گفته می‌شود را تولید می‌کنیم. MSE به‌صورت

$$\widehat{MSE}_m(\bar{g}_n) = \frac{1}{m} \sum_{k=1}^{m} (\bar{g}_n^{(k)} - \bar{g}^*)^2$$

برآورد می‌شود. مقادیر رابطه فوق اجازه می‌دهد تا تنها از نظر سازگاری، برآورد $E_\varpi g(X)$ را بررسی کنیم.

---

[۳۷] Autocorrelation Time

[۳۸] Mean Square Error

ما همچنین حرکت زنجیرها را در ناحیه فضای حالت با استفاده از مربع امید ریاضی فاصله اقلیدسی پرش[۳۹]، (ESEJD)، مقایسه می‌کنیم که بیان کننده مربع امید ریاضی فاصله متوالی بین زنجیر مارکوف $X^{(k)}$ و $X^{(k+1)}$ است. اگر $\|\cdot\|_2$ نشان‌دهنده نرم اقلیدسی استاندارد باشد آن‌گاه ESEJD امید ریاضی مقداری از میانگین مربع فاصله اقلیدسی پرش[۴۰] ، (MSEJD)، برای یک زنجیر به طول n، ثابت است

$$MSEJD = \frac{1}{n-1}\sum_{k=1}^{n-1}\|X^{(k+1)} - X^{(k)}\|_2^2 = ((E|X^{(k+1)} - X^{(k)}|^2)^{\frac{1}{2}})^2$$

با توجه به m تکرار مستقل از هر زنجیر، هر کدام با طول n، می‌توانیم ESEJD

$$\widehat{ESEJD}_m = \frac{1}{m}\sum_{k=1}^{m} MSEJD^{(k)}$$

برآورد کنیم.

سایر روش‌های بررسی همگرایی

همان‌طور که قبلا اشاره شد، در بسیاری از موارد تولید نمونه مستقیم مشکل است. در این موارد می‌توان با استفاده از یک زنجیر مارکوف، به‌طور غیر مستقیم تولید نمونه کرد. اما از آن‌جایی‌که نمونه تولید شده از زنجیر مارکوف وابسته است، کیفیت یک نمونه مستقل با حجمی مشابه را ندارد. کیفیت چنین نمونه‌ای را می‌توان از طریق کمیتی به نام حجم نمونه موثر[۴۱] (ESS) سنجید (گلمن و همکاران، ۱۹۹۸؛ رابرت و کسلا، ۲۰۰۴). هرچه مقدار ESS محاسبه شده برای هر پارامتر به تعداد نمونه‌های تولید شده نزدیک‌تر باشد، نمونه‌های تولید شده از کیفیت بالاتری برخوردار هستند.

بنابراین، بررسی همگرایی زنجیر به توزیع مانای خود، یکی از مباحث مهمی است که در روش نمونه‌گیری MCMC مطرح می‌شود. بحثی که در این‌جا مطرح می‌شود این است که چگونه تشخیص دهیم یک زنجیر همگراست. زیرا انتظار داریم زنجیر تولید شده در نهایت به توزیع مانایی که مدنظر است، همگرا شود. کاربران MCMC، همگرایی را با پیاده کردن روش‌های تشخیص همگرایی روی خروجی تولید شده از نمونه‌گیری، بررسی می‌کنند. روش‌های متعددی برای بررسی همگرایی زنجیرهای MCMC وجود دارند. از جمله این روش ها می‌توان به معیارهای همگرایی گلمن و روبین (۱۹۹۲)، جی‌وک (۱۹۹۲) و هیدل‌برگر و ولچ (۱۹۸۳) اشاره کرد.

۹.۱ مدل‌های آمیخته خطی تعمیم یافته

تقریبا در اغلب موارد استفاده از روش‌های آماری، مرکز توجه روی میانگین‌ها و تغییرات آن‌هاست، که منجر به ایجاد روش‌های استنباط درباره میانگین‌ها یا درباره ماهیت آن‌ها می‌شود. معمول‌ترین

[۳۹] Expected Square Euclidean Jump Distance

[۴۰] Mean Square Euclidean Jump Distance

[۴۱] Effective Sample Size

مدل رگرسیونی، مدل خطی[۴۲] (LM) است. در یک مدل خطی، با متغیر پاسخ Y و ماتریس طرح X، تابع رگرسیونی، که همان میانگین شرطی $Y|X$ است، بر حسب ترکیب خطی از X به‌صورت $E[Y|X] = X\beta$ در نظر گرفته می‌شود. در این مدل، ضرایب β بردارهای ثابت نامعلوم، یا به عبارتی اثرات ثابت[۴۳] نامیده می‌شوند.

پذیره اساسی در مدل خطی، استقلال بین مشاهدات است. اما در داده‌های طولی[۴۴] و سری زمانی بین داده‌ها وابستگی وجود دارد، و این وابستگی باید در مدل لحاظ شود، در این حالت، تعمیمی از این مدل‌ها معروف به مدل آمیخته خطی[۴۵] (LMMs) مورد استفاده قرار می‌گیرد. مدل‌های خطی تعمیم‌یافته (نلدر و ودربرن، ۱۹۷۲) رده‌ای از مدل‌ها هستند که برای مدل‌بندی پاسخ‌های غیرنرمالی که دارای توزیعی از خانواده توزیع‌های نمایی[۴۶] (یا شبیه به آن) هستند، معرفی شدند.

برای در نظر گرفتن وابستگی بین پاسخ‌ها، استفاده از متغیرهای پنهان[۴۷] یا همان اثرات تصادفی[۴۸] یک راه رایج و مناسب است. میانگین پاسخ در یک LMM به‌صورت $E[Y|X, Z, b] = X\beta + Zb$ نمایش داده می‌شود که در آن X ماتریس طرح متناظر با اثرات ثابت β و Z، ماتریس طرح متناظر با بردار اثرات تصادفی b می‌باشند (مک‌کالاک و سیرل، ۲۰۰۱). اثرات تصادفی معمولا به عنوان متغیرهای تصادفی نرمال (با میانگین صفر) تلقی می‌شوند و پارامترهایی که توزیع این اثرات را توصیف می‌کنند، برآورد می‌شوند.

ذات این تعمیم در این‌گونه مدل‌ها دو چیز است: اول، متغیرهای پاسخ می‌توانند غیر نرمال باشند؛ دوم، میانگین پاسخ لزوما به‌صورت ترکیبی خطی از متغیرهای تبیینی نیست، بلکه تابعی از آن به‌نام تابع پیوند[۴۹]، به‌صورت ترکیب خطی از متغیرهای تبیینی در نظر گرفته می‌شود.

در یک GLM[۵۰] ، اگر η رابطه‌ای خطی از متغیرهای تبیینی و ضرایب رگرسیونی، به‌نام پیشگوی خطی[۵۱] باشد، آن‌گاه $g(\mu) = \eta = X\beta$ که $g(\cdot)$ یک تابع پیوند است. این تابع با پیوند میانگین پاسخ و پیشگوی خطی، میانگین را مدل‌بندی می‌کند. از جمله توابع مورد استفاده در $GLMs$، توابع پیوند لجیت[۵۲]، $g(\mu) = ln\frac{\mu}{1-\mu}$، پروبیت[۵۳]، $g(\mu) = \Phi^{-1}(\mu)$ و لگاریتمی، $g(\mu) = ln(\mu)$ می‌باشند. برای آگاهی بیشتر این توابع به مک‌کالاک و نلدر (۱۹۸۹) مراجعه کنید.

---

[۴۲] Linear Model
[۴۳] Fixed effects
[۴۴] Longitudinal data
[۴۵] Linear mixed models
[۴۶] Exponential family distributions
[۴۷] Latent variables
[۴۸] Random effects
[۴۹] Link function
[۵۰] Generalized Linear Models
[۵۱] Linear preditor
[۵۲] Logit
[۵۳] Probit

فصل ۲

به‌روز رسانی مولفه به مولفه

در این فصل به دنبال معرفی الگوریتم‌های تولید نمونه یک متغیر یا زیربلوکی از متغیرها در هر مرحله به‌جای تولید هم‌زمان همه متغیرهای توزیع پسین هستیم. همچنین در کنار برتری‌های به‌روز رسانی به روش مولفه به مولفه، استراتژی‌های مختلف برای ترکیب مولفه‌ها را بیان می‌کنیم.

۱.۲ مقدمه

همان‌طور که اشاره کردیم، الگوریتم‌های MCMC مختلفی برای تقریب (تولید نمونه از) توزیع‌های هدف با بعد بالا معرفی شده‌اند. با این حال، به منظور به حداکثر رساندن کارایی نمونه‌های تولیدی به روش مونت کارلو، ضروری است تا الگوریتم‌هایی طراحی شوند که زنجیرهای مارکوف تولیدشده توسط آن‌ها به سرعت آمیخته شوند.

در بسیاری از مسایل واقعی، بعد متغیرهای توزیع پسین بالا بوده و در روش‌های MCMC، معمول است که به‌جای به‌روز کردن هم‌زمان همه متغیرها در هر مرحله، فقط یک متغیر یا زیربلوکی از متغیرها به‌روز شود که به آن به‌روز رسانی مولفه به مولفه می‌گویند (نیل و رابرتز، ۲۰۰۶).

نمونه‌گیر گیبز که در آن، در هر مرحله از توزیع شرطی کامل متغیرها نمونه‌گیری می‌شود، حالت خاصی از این نوع به‌روز رسانی است. الگوریتم متروپولیس–هستینگز درون گیبز (که همان الگوریتم گیبز است که برای تولید نمونه از برخی از توزیع‌های شرطی کامل از الگوریتم متروپولیس–هستینگز استفاده می‌شود) نیز، نسخه دیگری از این نوع به‌روز رسانی می‌باشد.

در مورد رفتارهای همگرایی، شامل ارگودیک یکنواخت و هندسی، صورت‌های مختلف زنجیرهای MCMC زمانی‌که همه متغیرها هم‌زمان به‌روز می‌شوند، پژوهش‌های متعدد زیادی انجام شده‌اند. به عنوان مثال، منگرسن و تووییدی (۱۹۹۶) و تیرنی (۱۹۹۴) به مطالعه الگوریتم MHIS پراختند؛ منگرسن و تووییدی (۱۹۹۶)، رابرتز و تووییدی (۱۹۹۶)، یارنر و هنسن (۲۰۰۰)، کریستنسن و همکاران (۲۰۰۱) و جانسون و گیر (۲۰۱۲) به ارایه شرایطی پرداختند که تحت آن‌ها زنجیر مارکوف یک الگوریتم متروپولیس، ارگودیک هندسی باشد. همچنین مین و تووییدی (۱۹۹۴)، گیر (۱۹۹۹)، و یارنر و هنسن (۲۰۰۰)، به

مطالعه نرخ همگرایی الگوریتم‌های MH پرداختند. اما هیچ‌کدام از نتایج آن‌ها برای زنجیرهای مولفه به مولفه MH به‌کار گرفته نشده‌اند. تنها روش مولفه به مولفه‌ای که کمی مورد توجه قرار گرفته است، نمونه‌گیر گیبز است. برای مثال می‌توانید داس و هابرت (۲۰۱۲) و جانسون و جونز (۲۰۱۰) را مشاهده کنید.

برای معرفی یک الگوریتم مولفه به مولفه در حالت کلی و توسعه استراتژی‌های ترکیب، ابتدا باید نمادهای لازم را معرفی کنیم. فرض کنید ϖ یک توزیع احتمال با تکیه‌گاه $X \subset \mathbb{R}^q$، $q \geq 1$، باشد. برای بیان به‌تر و منسجم روش‌های مولفه به مولفه، به‌طور خیلی خلاصه، الگوریتم پایه‌ای متروپولیس-هستینگز را (علی‌رغم آن‌که در فصل اول به‌طور کامل تشریح شده است)، بار دیگر معرفی می‌کنیم. فرض کنید ϖ دارای تابع چگالی $\pi(\cdot)$ و $X^{(k)} = x$ مقدار جاری زنجیر باشد. همچنین فرض کنید $p(\cdot,\cdot)$ تابع چگالی پیشنهادی باشد. در الگوریتم MH حالت به‌روزشده $X^{(k+1)}$ به صورت زیر به‌دست می‌آید:

۱) تولید y از چگالی پیشنهادی $p(y|x)$.

۲) محاسبه احتمال پذیرش $\alpha(x,y)$ که در آن

$$\alpha(x,y) = \min\left\{1, \frac{\pi(y)p(x|y)}{\pi(x)p(y|x)}\right\}.$$

۳) قرار دادن $X^{(k+1)} = y$ با احتمال $\alpha(x,y)$ و قرار دادن $X^{(k+1)} = x$ با احتمال $1 - \alpha(x,y)$.

بنابراین ساخت یک الگوریتم نمونه‌گیری MH به انتخاب تابع چگالی پیشنهادی $p(\cdot,\cdot)$ خلاصه می‌شود. اگر $p(x,y) = p(y,x)$، الگوریتم حاصل را متروپولیس می‌گویند. افزون بر این، اگر برای همه مقادیر ممکن x و y، $p(x,y) = p(x-y) = p(y-x)$، الگوریتم حاصل یک الگوریتم MHRW است. چنان‌چه چگالی پیشنهادی نسبت به حالت جاری زنجیر مستقل باشد، یعنی $p(y,x) = p(y)$، نتیجه الگوریتم MHIS خواهد بود.

انتخاب تابع چگالی پیشنهادی، به‌ویژه زمانی که بعد q بزرگ است یا تکیه‌گاه ϖ پیچیده است، می‌تواند چالش‌برانگیز باشد. این چالش، آماردانان این حوزه را بر آن داشته است که در مورد مقیاس‌بندی بهینه الگوریتم‌های MH به پژوهش بپردازند. بخشی از این پژوهش‌ها منجر به پیدایش الگوریتم‌های MCMC سازوار[1] شده است. این نوع الگوریتم‌ها اجازه می‌دهند هسته تابع چگالی پیشنهادی در طی فرآیند شبیه‌سازی MCMC، برای آمیختگی به‌تر زنجیر، تغییر کند (بدراد و روزنتال، ۲۰۰۸؛ روزنتال، ۲۰۱۱).

بخش دیگر پژوهش‌های مذکور به معرفی روش‌های مولفه به مولفه منجر شده است. در روش‌های مولفه به مولفه، بردار متغیرهای x به d زیربلوک $x = (x_1, \ldots, x_d)$ تجزیه می‌شود، که در آن $d \leq q$. بعد هر کدام از این زیربلوک‌ها ۱ یا بزرگتر از ۱ است، به‌طوری که مجموع بعدهای d زیربلوک برابر q خواهد بود. ساختار کلی یک الگوریتم مولفه به مولفه در رده الگوریتم‌های MCMC به‌صورت زیر خلاصه می‌شود:

۱) با شرط داشتن مقدار جاری زنجیر، ابتدا زیربلوک اول، به شرط ثابت نگه داشتن سایر زیربلوک‌ها، با یکی از روش‌های MCMC به‌روز می‌شود.

[1] Adaptive MCMC algorithms

۲) زیربلوک دوم، به شرط ثابت بودن زیربلوک اول در مقدار به‌روزشده مرحله ۱) و سایر زیربلوک‌ها در مقدار جاری، با یکی از روش‌های MCMC به‌روز می‌شود.

۳) به‌طور مشابه زیربلوک‌های سوم تا d ام به‌روز می‌شوند.

این مراحل به تعداد دلخواه به اندازه کافی بزرگ، برای به‌دست آوردن یک نمونه از توزیع پسین، تکرار می‌شوند. دقت کنید اگر $d = ۱$، آن‌گاه روش به‌روز کردن همان روش هم‌زمان همه متغیرها خواهد بود و در حالتی که $d = q$، در هر مرحله فقط یک متغیر به‌روز می‌شود.

به‌طور کلی، انتخاب به‌روز رسانی توام یا مولفه به مولفه چندان واضح نیست (رابرتز و ساهو، ۱۹۹۷). اما می‌توان یک قاعده کلی را مطرح کرد:

> اگر مولفه‌های ϖ همبستگی ضعیفی داشته باشند، به‌روز رسانی هم‌زمان همه آن‌ها ممکن است کارا نباشد.

به‌عنوان مثال، نیل و رابرتز (۲۰۰۶) با در نظر گرفتن توزیع‌های هدفی که مولفه‌های i.i.d. دارند، نشان دادند که نسخه ایده‌آل یک الگوریتم MCMC مولفه به مولفه، الگوریتمی است که در هر زمان تنها یک مولفه را به‌روز کند. آن‌ها همچنین نشان دادند برای استفاده از الگوریتم‌های MH تعدیل‌شده با لانگوین[۲] (رابرتز و روزنتال، ۱۹۹۷) به‌روز رسانی بلوکی کاراتر است از توام. اما در یک نگاه کلی، می‌توان دلایل برتری روش‌های مولفه به مولفه را (جاهایی که برتر است) به دو مورد زیر خلاصه کرد:

۱) در مسایلی که بعد فضای حالت بالا و/یا تکیه‌گاه ϖ پیچیده است، استفاده از الگوریتم‌های مولفه به مولفه کاراتر است؛ زیرا اگر بعد بالا باشد باید یک تابع چگالی پیشنهادی با این بعد انتخاب شود که می‌تواند سخت باشد. از طرفی، چگالی پیشنهادی منتخب باید در تکیه‌گاه ϖ قرار گیرد یا آن را پوشش دهد. بنابراین به دلیل کاربردهای وسیع روش‌های MCMC برای کاربران، و به دلیل این‌که به‌روز رسانی با بعد بالا دارای محاسبات پیچیده‌تر و بیش‌تری است، به‌روز رسانی با بعد پایین‌تر ترجیح داده می‌شود. در نتیجه شکستن مساله شبیه‌سازی با بعد بالا به چندین مساله با بعد پایین، به روش مولفه به مولفه، می‌تواند انتخاب تابع چگالی پیشنهادی را ساده‌تر کند.

۲) وابستگی ضعیفی که بین مولفه‌های ϖ وجود دارد، باعث می‌شود که الگوریتم به‌روز رسانی توام همه آن‌ها به کندی همگرا شود. بنابراین با بلوک‌بندی مولفه‌ها، به‌طوری که مولفه‌های درونی بلوک‌ها وابستگی بالایی داشته و خود بلوک‌ها تقریبا از هم مستقل باشند، می‌توان به همگرایی الگوریتم کلی سرعت و کارایی قابل ملاحظه‌ای بخشید.

سوال اساسی که در روش‌های مولفه به مولفه مطرح می‌شود، نحوه ترکیب کردن مولفه‌های به‌روزشده است. برای پاسخ به این سوال، استراتژی‌های مختلفی برای ترتیب و فراوانی به‌روز رسانی مولفه‌ها توسط افراد مختلف پیشنهاد شده‌اند. از جمله آن‌ها می‌توان به اتصال همگی یا انتخاب یک دنباله تصادفی به‌عنوان حالت زنجیر توام اشاره کرد (فورت و همکاران، ۲۰۰۳). انواع متفاوت استراتژی‌های ترکیب در دو رده کلی ترکیب[۳] و آمیختن[۴] دسته‌بندی می‌شوند. این روش‌های ترکیب کردن باعث اجرای ساده‌تر و عملکرد

[۲] Metropolis-Hastings Adjusted Langevin algorithm

[۳] Composition

به‌تر الگوریتم MCMC نسبت به روش‌های به‌روز رسانی هم‌زمان همه متغیرها می‌شوند. در بخش بعدی به معرفی دو رده کلی استراتژی‌های ترکیب و برخی ویژگی‌های آن‌ها می‌پردازیم.

۲.۲ استراتژی‌های ترکیب

انتخاب استراتژی به‌روز رسانی برای نمونه‌گیری از توزیع هدف (پسین)، می‌تواند روی سرعت همگرایی به‌طور چشم‌گیری اثرگذار باشد. دو رده کلی استراتژی‌ها برای ترکیب هسته‌های مارکوف، آمیختگی و ترکیب هستند. ابتدا این دو استراتژی کلی را تعریف می‌کنیم.

فرض کنید $P_1, \dots, P_d$ هسته‌های مارکوفی باشند که همگی دارای توزیع مانای ϖ هستند. همچنین فرض کنید

$$\mathbb{P}^d = \left\{ (r_1, \dots, r_d) \in \mathbb{R}^d : r_i > 0, \ \sum_{i=1}^{d} r_i = 1 \right\}$$

که در آن r_i احتمال انتخاب به‌روز رسانی i امین مولفه است و $r_i > 0$ به این معناست که هر مولفه اغلب بی‌نهایت بار دیده خواهد شد. اکنون دو استراتژی مذکور را به صورت زیر تعریف می‌کنیم.

تعریف ۱.۲.۲. (استراتژی ترکیب). هسته ترکیب به صورت

$$P_{comp}(x, .) = (P_1, \dots, P_d)(x, .)$$

تعریف می‌شود.

تعریف ۲.۲.۲. (استراتژی آمیختگی). هسته آمیختگی به صورت

$$P_{mix}(x, .) = r_1 P_1(x, .) + \cdots + r_d P_d(x, .), \quad r = (r_1, \dots, r_d) \in \mathbb{P}^d$$

تعریف می‌شود. معمول است که می‌گویند P_{mix} احتمال‌های انتخاب r را دارد.

اولین نتیجه‌ای که می‌توان در مورد این دو رده استراتژی کلی گفت، در قضیه زیر ارایه شده است. این قضیه بیان می‌کند توزیع مانای هسته‌های مارکوف حاصل از دو استراتژی ترکیب و آمیختن همان توزیع مانای هسته‌های مارکوف به‌کار گرفته شده برای ساخت الگوریتم‌های مولفه به مولفه است.

قضیه ۳.۲.۲. *دو هسته P_{mix} و P_{comp} دارای توزیع مانای ϖ هستند.*

برهان. باید نشان دهیم $\varpi P_{comp} = \varpi$ و $\varpi P_{mix} = \varpi$. ابتدا برهان را برای هسته P_{comp} می‌نویسیم. داریم

$$\varpi P_{comp} = \varpi (P_1, \dots, P_d)(x, .)$$

چون هر P_i مولفه i ام (مولفه متناظر با اندیس خود) را به‌روز رسانی می‌کند و چون $P_1, \dots, P_d$ هسته‌های مارکوف با توزیع مانای ϖ هستند، برابری

$$\varpi (P_1, \dots, P_d)(x, .) \ = \varpi$$

[۴] Mixing

به سادگی نتیجه می‌شود. برای هسته P_{mix} داریم

$$\begin{aligned}\varpi P_{mix} &= r_1 \varpi P_1(x,\cdot) + r_2 \varpi P_2(x,\cdot) + \cdots + r_d \varpi P_d(x,\cdot) \\ &= r_1 \varpi + r_2 \varpi + \cdots + r_d \varpi \\ &= \varpi(\sum_{i=1}^{d} r_i) = \varpi\end{aligned}$$

و برهان کامل است. □

از این دو استراتژی می‌توان برای تولید بسیاری از الگوریتم‌های مولفه به مولفه استفاده کرد. در بخش بعدی به حالت‌های خاصی از آن‌ها که منجر به این الگوریتم‌ها می‌شوند، می‌پردازیم.

۳.۲ استراتژی‌های معمول ترکیب

فرض کنید π یک تابع چگالی از توزیع ϖ نسبت به اندازه $\mu = \mu_1 \times \cdots \times \mu_d$ باشد، به‌طوری‌که تکیه‌گاه آن $X = X_1 \times \cdots \times X_d \subseteq \mathcal{B}$ است، که در آن $\mathcal{B}$ سیگما جبر بورل است و برای هر $i = 1, \ldots, d$، $X_i \subseteq \mathbb{R}^{b_i}$. دقت کنید $b_1 + \cdots + b_d = q$ است. همچنین فرض کنید برای $x \in \chi$ مجموعه $x_{(i)} = x \backslash x_i$ نشان‌دهنده بردار x به‌جز x_i است و برای $i = 1, \ldots, d$، تابع چگالی $g_i(y_i|x)$ در تعریف

$$\pi(y_i|x_{(i)}) = \int g_i(y_i|x)\pi(x_i|x_{(i)})\mu_i(dx_i) \qquad (۱.۲)$$

صدق کند. این به آن معنی است که چگالی شرطی $\pi(x_i|x_{(i)})$ برای $g_i(y_i|x)$ ماناست. توجه داشته باشید که اگر $g_i(y_i|x)$ تابع چگالی پیشنهادی متناظر با الگوریتم گیبز باشد، یعنی

$$g_i(y_i|x) = \pi(y_i|x_{(i)})$$

آن‌گاه رابطه (۱.۲) برقرار است. همچنین اگر $g_i(y_i|x)$ تابع چگالی پیشنهادی متناظر با الگوریتم MH باشد که توزیع مانای آن $\pi(y_i|x_{(i)})$ است، آن‌گاه شرط (۱.۲) باز هم برقرار است.

تحت (۱.۲)، هسته مارکوف P_i را به صورت

$$P_i(x, A) = \int_A g_i(y_i|x)\delta(y_{(i)} - x_{(i)})\mu(dy) \qquad (۲.۲)$$

تعریف می‌کنیم که در آن $A \in \mathcal{B}$ و δ، دلتای دیراک[۵] تعمیم‌یافته است که به صورت

$$\delta(t) = \begin{cases} \infty & t = 0 \\ 0 & o.w \end{cases}$$

تعریف می‌شود. مهم‌ترین ویژگی تابع دلتای دیراک تعمیم‌یافته آنست که به ازای هر تابع اندازه‌پذیر $f(t)$ می‌توان نوشت

$$\int dt f(t)\delta(t) = f(0).$$

[۵] Dirac's delta

بنابراین $\delta(t)$ به‌جز جاهایی که $t = ۰$ باشد، دیده نمی‌شود. در نتیجه مهم نیست که تابع $f(t)$ چه مقادیری در سایر نقاط به‌جز صفر به‌دست می‌آورد و می‌توان نوشت $f(t)\delta(t) = f(۰)\delta(t)$ یا به‌عبارت دیگر

$$\int f(t)\delta(t - t_۰)dt = f(t_۰).$$

با تعریف هسته مارکوف P_i در رابطه (۲.۲)، قضیه زیر قابل طرح و اثبات است.

قضیه ۱.۳.۲. *هسته مارکوف تعریف‌شده در (۲.۲) دارای توزیع مانای ϖ است.*

برهان. باید نشان دهیم $\varpi P_i = \varpi$. داریم

$$\begin{aligned}\int_X \pi(x)g_i(y_i|x)\delta(y_{(i)} - x_{(i)})\mu(dx) &= \int_X \pi(x_{(i)}, x_i)g_i(y_i|x_{(i)}, x_i)\delta(y_{(i)} - x_{(i)})\mu(dx)\\ &= \int_{x_i}\int_{x_{(i)}} \pi(x_i|x_{(i)})\pi(x_{(i)})g_i(y_i|x_{(i)}, x_i)\delta(y_{(i)} - x_{(i)})\mu_{(i)}(dx_{(i)})\mu_i(dx_i)\\ &= \int_{x_{(i)}} \pi(x_i|y_{(i)})\pi(y_{(i)})g_i(y_i|y_{(i)}, x_i)\mu_i(dx_i)\\ &= \int_{x_i} \pi(y_i, y_{(i)}, x_i)\mu_i(dx_i) = \pi(y)\end{aligned}$$

و برهان کامل است. □

از آن‌جایی که به‌روز رسانی‌های مولفه به مولفه به صورت بلوکی و مستقل به‌روز می‌شوند و امکان جابه‌جایی هر مولفه به هر ناحیه از فضای حالت وجود ندارد، پس ϖ-تحویل‌ناپذیر نیستند. از طرفی، الگوریتم‌های MCMC مفید باید ویژگی ϖ-تحویل‌ناپذیری را داشته باشند. بنابراین برای دسترسی به چنین الگوریتم‌هایی باید P_i ها را به‌گونه‌ای ترکیب کنیم که دارای این ویژگی باشند. نحوه ترکیب کردن P_i ها منجر به معرفی استراتژی‌های ترکیب مختلف می‌شود که در ادامه سه استراتژی معروف و پر کاربرد را معرفی می‌کنیم.

استراتژی اسکن تصادفی

فرض کنید $r \in \mathbb{P}^d$، احتمال‌های انتخاب متناظر با d مولفه باشد. در این صورت، می‌توان هسته مارکوف اسکن تصادفی[۶] (RS) را به صورت

$$P_{RS}(x, A) = \sum_{i=۱}^{d} r_i P_i(x, A) \qquad (۳.۲)$$

تعریف کرد، که چگالی انتقال مارکوف آن به صورت

$$h_{RS}(y|x) = \sum_{i=۱}^{d} r_i g_i(y_i|x)\delta(y_{(i)} - x_{(i)})$$

خواهد بود. قضیه زیر بیان می‌کند توزیع مانای حاصل از هسته مارکوف این نوع استراتژی، همان توزیع مطلوب ϖ است.

[۶] Random scan

قضیه ۲.۳.۲. *هسته مارکوف روش اسکن تصادفی، P_{RS}، دارای توزیع مانای ϖ است.*

برهان. باید نشان دهیم $\varpi P_{RS} = \varpi$. داریم

$$\int_X \pi(x) \sum_{i=1}^{d} r_i \int_A g_i(y_i|x)\delta(y_{(i)} - x_{(i)})\mu(dx) = \int_X \sum_{i=1}^{d} r_i \int_A \pi(x) g_i(y_i|x)\delta(y_{(i)} - x_{(i)})\mu(dx)$$
$$= \int_X \sum_{i=1}^{d} r_i \pi(y) = \sum_{i=1}^{d} r_i \pi(y) = \pi(y) \sum_{i=1}^{d} r_i = \pi(y)$$

و برهان کامل است. □

در این روش، در هر مرحله، یکی از مولفه‌ها با احتمال انتخاب متناظرش به‌روز می‌شود، در حالی‌که سایر $d-1$ مولفه ثابت باقی می‌مانند. برای درک بهتر این استراتژی، الگوریتم نمونه‌گیری گیبز تحت آن را به صورت زیر معرفی می‌کنیم:

۱) تعیین احتمال‌های انتخاب $r = (r_1, r_2, \ldots, r_d)$ و مقدار اولیه $X^{(0)}$. سپس برای $t = 1, 2, \ldots$ تکرار مرحله بعدی.

۲) تولید $X^{(t)} = (X_1^{(t)}, X_2^{(t)}, \ldots, X_d^{(t)})$ در t امین تکرار، به کمک مراحل زیر

- انتخاب i از مجموعه $\{1, 2, \ldots, d\}$ با احتمال r_i.
- تولید

$$X_i^{(t)} \sim \pi(X_i|X_{(i)}^{(t-1)})$$

و قرار دادن

$$X^{(t)} = \left(X_1^{(t-1)}, \ldots, X_{i-1}^{(t-1)}, X_i^{(t)}, X_{i+1}^{(t-1)} \ldots, X_d^{(t-1)}\right)$$

استراتژی ترکیب

راه دوم برای آمیختن P_i ها، ترکیب[۷] (C) است. در این روش، مولفه‌های X با یک نظم و ترتیب ثابت (غیرتصادفی) و از پیش تعیین‌شده به‌روز می‌شوند. بدون از دست دادن کلیت، می‌توان پذیرفت ترتیب ترکیب همان $(1, 2, \ldots, d)$ باشد. در نتیجه هسته مارکوف به صورت

$$P_C(x, A) = (P_1 \ldots P_d)(x, A) \qquad (۴.۲)$$

تعریف می‌شود، به‌طوری که تابع چگالی انتقال مارکوف آن به شکل

$$h_C(y|x) = g_1(y_1|x) g_2(y_2|y_1, x_{(1)}) \ldots g_d(y_d|y_{(d)}, x_d)$$

است.

[۷] Combine

نتیجه ۳.۳.۲. *مشابه قضیه ۳.۲.۲، می‌توان نشان داد $\varpi P_C = \varpi$. یعنی توزیع مانای هسته مارکوف حاصل از این استراتژی نیز توزیع مطلوب ϖ است.*

مشابه استراتژی اسکن تصادفی، برای درک به‌تر استراتژی ترکیب، الگوریتم نمونه‌گیر گیبز تحت این نوع استراتژی را معرفی می‌کنیم:

۱) تعیین مقدار اولیه $X^{(۰)}$ و تکرار مرحله بعدی برای $t = ۱, ۲, \ldots$.

۲) تولید $X^{(t)} = (X_۱^{(t)}, X_۲^{(t)}, \ldots, X_d^{(t)})$ در t امین تکرار با اجرای مراحل زیر:

- تولید

$$X_۱^{(t)} \sim \pi(X_۱ | X_{(۱)}^{(t-۱)})$$
$$X_۲^{(t)} \sim \pi\left(X_۲ | X_۱^{(t)}, X_{(۱,۲)}^{(t-۱)}\right)$$
$$\vdots$$
$$X_d^{(t)} \sim \pi\left(X_d | X_{(d)}^{(t)}\right).$$

استراتژی دنباله اسکن تصادفی

در استراتژی ترکیب با d مولفه، تعداد ترتیب‌های ممکن برای به‌کار گرفتن برابر $d!$ است. در استراتژی ترکیب تنها یکی از این ترتیب‌ها از قبل تعیین می‌شود و به‌روز رسانی مولفه‌ها بر اساس آن ترتیب انجام می‌شود. اما برای در نظر گرفتن اثر سایر ترتیب‌ها، یک راه طبیعی استفاده از آمیختن همه هسته‌های مارکوف حاصل از ترتیب‌ها می‌باشد. اما اگر d (یعنی تعداد بلوک‌ها) متوسط یا بزرگ باشد، $d!$ خیلی بزرگ خواهد شد و اجرای این استراتژی کار ساده و کم‌هزینه‌ای نیست و ما را از هدف اصلی، یعنی کارایی محاسباتی الگوریتم MCMC، دور می‌کند.

راهی برای دور شدن از این مشکل، استفاده از برخی از این ترتیب‌ها، مثلا p تا، است به‌طوری که $p < d!$. به این استراتژی، آمیختگی دنباله‌ای[۸] می‌گویند که نماد اختصاری آن (RQ)[۹] است. اگر $r \in \mathbb{P}^p$، هسته آمیختگی دنباله‌ای به صورت

$$P_{RQ}(x, A) = \sum_{j=۱}^{p} r_j P_{C,j}(x, A) \qquad (۵.۲)$$

تعریف می‌شود، که در آن $P_{C,j}$ هسته‌های مارکوف ساخته‌شده به روش ترکیب ولی با ترتیب‌های متفاوت منتخب هستند. از طرفی، تابع چگالی انتقال مارکوف P_{RQ} به شکل

$$h_{RQ}(y|x) = \sum_{j=۱}^{p} r_j h_{C,j}(y|x)$$

خواهد بود.

[۸] Sequence mixing

[۹] Random sequence

نتیجه ۴.۳.۲. *با توجه به قضیه ۳.۲.۲ و نتیجه ۳.۳.۲، چون برای هر j هسته $P_{C,j}$ دارای توزیع مانای ϖ است، توزیع مانای هسته مارکوف P_{RQ} نیز ϖ می‌باشد؛ یعنی $\varpi P_{RQ} = \varpi$.*

برای اجرای این روش، ابتدا p (که معمولا کوچک اختیار می‌شود) تعیین می‌شود. در استراتژی آمیختگی دنباله‌ای، مانند استراتژی ترکیب، تمام d مولفه در هر تکرار، بر اساس ترتیب مشخص‌شده در آن تکرار، به‌روز می‌شوند. اگر ترتیب به‌روز رسانی i ام را با نماد $(i(۱), \ldots, i(d))$ نشان دهیم که در آن $i(j) \in \{۱, \ldots, d\}$ و مجموعه احتمال‌های p ترتیب (جایگشت) را با $q = \{q_۱, q_۲, \ldots, q_p\}$ بیان کنیم، به‌طوری که $\sum_{i=۱}^{p} q_i = ۱$، آن‌گاه در هر تکرار، ترتیب به‌روز رسانی به‌طور تصادفی و با توجه به مجموعه q انتخاب می‌شود. الگوریتم نمونه‌گیر گیبز تحت استراتژی آمیختگی دنباله‌ای به صورت زیر است:

۱) تعیین احتمال‌های جایگشت $q = \{q_۱, q_۲, \ldots, q_p\}$ و مقدار اولیه $X^{(۰)}$. سپس تکرار مرحله بعدی برای $t \geq ۱$.

۲) تولید $X^{(t)} = (X_۱^{(t)}, X_۲^{(t)}, \ldots, X_d^{(t)})$ در t امین تکرار با اجرای مراحل زیر:

- انتخاب تصادفی یک اندیس $i \in \{۱, ۲, \ldots, d\}$ با احتمال q_i.
- تولید $X^{(t)}$ بر اساس i امین ترتیب $(i(۱), i(۲), \ldots, i(d))$:

$$X_{i(۱)}^{(t)} \sim \pi\left(X_{i(۱)} | X_{(i(۱))}^{(t-۱)}\right)$$

$$X_{i(۲)}^{(t)} \sim \pi\left(X_{i(۲)} | X_{i(۱)}^{(t)}, X_{(i(۱),i(۲))}^{(t-۱)}\right)$$

$$\vdots$$

$$X_{i(d)}^{(t)} \sim \pi\left(X_{i(d)} | X_{(i(d))}^{(t)}\right).$$

توجه داشته باشید که هسته‌های تعریف‌شده در روابط (۳.۲)، (۴.۲) و (۵.۲) نمونه‌های خاصی از استراتژی‌های کلی P_{mix} و P_{comp} هستند و همگی در شرط (۲.۲) صدق می‌کنند. نرخ همگرایی الگوریتم‌های مولفه به مولفه، هم به توزیع هدف وابسته است و هم به استراتژی‌های ترکیب کردن مولفه‌ها. این استراتژی‌ها می‌توانند زنجیرهایی با رفتارهای مجانبی مختلف نیز تولید کنند.

هر چند استراتژی ترکیب به‌طور گسترده استفاده می‌شود و برای کاربران بسیار آشنا است، اما مزایای ویژه‌ای در استفاده از دو استراتژی اسکن تصادفی و آمیختگی دنباله‌ای وجود دارند. برای مثال، روش اسکن تصادفی با توجه به ϖ-برگشت‌پذیر بودن، تحت محدودیت‌هایی، باعث ضعیف‌تر کردن (ساده کردن) شرایط برقراری یک CLT می‌شود. تاثیر انواع استراتژی‌های ترکیب کردن را، برای دست‌یابی به همگرایی یکنواخت و به‌ویژه هندسی، در فصل بعد بررسی می‌کنیم.

فصل ۳

نرخ همگرایی تحت به‌روز رسانی مولفه به مولفه

در این فصل نرخ‌های همگرایی روش‌های مولفه به مولفه را مطالعه می‌کنیم. در این راستا شرایطی که تحت آن برخی زنجیرهای مارکوف مولفه به مولفه به یک توزیع ثابت با نرخ هندسی همگرا می‌شوند را ارایه می‌دهیم و به ارتباط بین نرخ‌های همگرایی و استراتژی‌های مولفه به مولفه گوناگون می‌پردازیم. همچنان شرایط را برای ارگودیک یکنواخت مولفه به مولفه، الگوریتم متروپولیس- هستینگز با توزیع‌های پیشنهادی حالت مستقل، توسعه می‌دهیم. تحقیقات زیادی در خصوص ایجاد ارگودیک هندسی و یکنواخت برای نسخه‌های مختلف متروپولیس-هستینگز زمانی‌که تمام متغیرها به‌روز می‌شوند صورت گرفته است. برای بررسی جزییات بیش‌تر نمونه‌گیری متروپولیس-هستینگز مستقل، می‌توانید به منگرسن و توییدی (۱۹۹۶) و تیرنی (۱۹۹۴) مراجعه کنید و از طرفی جانسون و گیر (۲۰۱۲)، منگرسن و توییدی (۱۹۹۶)، رابرتز و توییدی (۱۹۹۶)، یارنر و هنسن (۲۰۰۰)، کریستنسن، مولر و ویچ‌پیترسن (۲۰۰۱) شرایطی که تحت آن، یک زنجیر متروپولیس-هستینگز ارگودیک هندسی باشد، بررسی کردند. تحقیقات دیگر، برای برقراری نرخ‌های همگرایی زنجیرهای متروپولیس، شامل یارنر و هنسن (۲۰۰۰)، گیر (۱۹۹۹) و مین و توییدی (۱۹۹۴) است.

توجه داشته باشید که هیچ یک از این نتایج نرخ همگرایی برای اجرای مولفه به مولفه متروپولیس-هستینگز به‌کار برده نشده است. بنابراین در الگوریتم به‌روز رسانی همه متغیرها ممکن است شرایط وجود ارگودیک هندسی برقرار نباشد.

۱.۳ مقدمه

اثبات همگرایی زنجیرهای مولفه به مولفه (به همراه استراتژی‌های مختلف ترکیب) با یک نرخ همگرایی هندسی می‌تواند مهم باشد زیرا برای کاربران آن‌ها این اطمینان را حاصل می‌کند که می‌توانند از نمونه‌های تولید شده از چنین زنجیرهایی با عنوان نمونه‌های با کیفیت، نزدیک به نمونه‌های مستقل، هم‌توزیع و

یکسان استفاده کنند (فلگال و جونز، ۲۰۱۰) . در این پایان‌نامه شرایطی را نمایش می‌دهیم که تحت آن‌ها برخی از زنجیرهای مارکوف مولفه به مولفه با یک نرخ هندسی به توزیع مانای خود همگرا شوند.

با این حال با یک شرح مختصری از برخی تکنیک‌ها برای وجود M و t در برقراری رابطه (۱.۱) شروع می‌کنیم. می‌توانید برای این‌که یک تعریفی داشته باشید به فصل ۱۵ از مین و توییدی (۱۹۹۳) و برای جزییات بیشتر به رابرتز و روزنتال (۲۰۰۴) و جونز و هابرت (۲۰۰۱) مراجعه کنید.

ماتریس احتمال انتقال، برای هر $x \in \chi$ و $A \in \mathcal{B}$ و $n \in \mathbb{Z}^+$ را به‌خاطر بیاورید، که n امین مرحله هسته انتقال زنجیر مارکوف، به‌صورت

$$P^n(x, A) = P(X^{(n+j)} \in A | X^{(j)} = x).$$

تعریف می‌شود.

گزاره ۱.۱.۳. *فرض کنید یک عدد صحیح مثبت n_0 و $\epsilon > 0$ و مجموعه $C \in \mathcal{B}$ و اندازه احتمال Q در $\mathcal{B}$ وجود دارد، به‌طوری‌که*

$$P^{n_0}(x, A) \geq \epsilon Q(A) \quad \text{برای همه} \quad x \in C, A \in \mathcal{B} \tag{۱.۳}$$

نتیجه ۲.۱.۳. *آن‌گاه شرط مینیمم‌سازی، برای مجموعه C، که کوچک‌ترین مجموعه نامیده می‌شود، برقرار است. ارگودیک یکنواخت با وجود شرط مینیمم‌سازی در χ معادل است.*

گزاره ۳.۱.۳. *فرض کنید*

$$PV(x) := E[V(X^{(t+1)}) | X^{(t)} = x] \tag{۲.۳}$$

اگر تابع $V : X \longrightarrow [1, \infty)$ و ثابت $0 < \gamma < 1$ و $K < \infty$ و مجموعه C وجود داشته باشد شرط رانش برقرار است به‌طوری‌که

$$PV(x) \leq \gamma V(x) + kI_C(x) \quad \text{برای همه} \quad x \in \chi. \tag{۳.۳}$$

نتیجه ۴.۱.۳. *اگر C کوچک باشد، آن‌گاه رابطه (۳.۳) با ارگودیک هندسی معادل است (مین و توییدی، ۱۹۹۳؛ رابرتز و روزنتال، ۱۹۹۷ و رابرتز و روزنتال، ۲۰۰۴).*

۲.۳ ارگودیک یکنواخت تحت آمیختگی و ترکیب

در این بخش برخی نتایج مربوط به نمونه‌گیری‌های P_{mix} و P_{comp} را در ابتدا بیان می‌کنیم و سپس به برخی از نمونه‌های مولفه به مولفه به‌طور تخصصی خواهیم پرداخت.

قضیه ۱.۲.۳. *اگر P_{mix} ارگودیک یکنواخت برای برخی احتمالات انتخاب باشد آن‌گاه آن ارگودیک یکنواخت برای همه احتمالات انتخاب است.*

برهان. فرض کنید $P_{mix,r}$، نشان‌دهنده P_{mix} برای انتخاب خاصی از $r \in \mathbb{P}^d$ باشد. همچنین فرض کنید $r \in \mathbb{P}^d$ ثابت و $t \in \mathbb{P}^d$ اختیاری باشد. توجه کنید که برای همه $i = 1, \ldots, d$ به‌طوری‌که $t_i > ar_i$ باشد، یک $a \in (0, 1)$ وجود دارد. بنابراین

$$P_{mix,t} = aP_{mix,r} + (1 - a)P_{mix,q}$$

که $q_i = (t_i - ar_i)/(۱ - a)$. با فرض این‌که $P_{mix,r}$ ارگودیک یکنواخت است از این‌رو یک مقدار صحیح مثبت $n_۰, \epsilon > ۰$ و اندازه احتمال Q وجود دارد، به‌طوری‌که با توجه به رابطه (۱.۳) داریم

$$P^{n_۰}_{mix,t}(x, A) \geq a^n_۰ P^{n_۰}_{mix,r}(x, A) \geq a^n_۰ \, \epsilon \, Q(A)$$

آن‌گاه شرط مینیمم‌سازی برقرار است و ارگودیک یکنواخت بودن همه احتمالات انتخاب ثابت می‌شود. □

بنابراین به راحتی دیده می‌شود که اگر یکی از مولفه‌های نمونه‌گیری، ارگودیک یکنواخت باشد آن‌گاه P_{mix} ارگودیک یکنواخت خواهد بود. همچنین کافی است که برای ایجاد ارگودیک یکنواخت P_{mix} ، P_{comp} را مطالعه کنید.

قضیه ۲.۲.۳. *فرض کنید P_{comp} ارگودیک یکنواخت باشد، آن‌گاه P_{mix} برای هر احتمال انتخاب، ارگودیک یکنواخت است.*

برهان. با فرض این‌که P_{comp} ارگودیک یکنواخت است از این‌رو، یک مقدار صحیح مثبت $n_۰$، ثابت ϵ و اندازه احتمال Q وجود دارد به‌طوری‌که بر اساس رابطه (۱.۳) داریم

$$\begin{aligned}
P^{n_۰ d}_{mix}(x, A) &= \int P^{n_۰ d-۱}_{mix}(y^{(۱)}, A) P_{mix}(x, dy^{(۱)}) \\
&\geq r_۱ \int P^{n_۰ d-۱}_{mix}(y^{(۱)}, A) P_۱(x, dy^{(۱)}) \\
&= r_۱ \int \int P^{n_۰ d-۲}_{mix}(y^{(۲)}, A) P_{mix}(y^{(۱)}, dy^{(۲)}) P_۱(x, dy^{(۱)}) \\
&\geq r_۱ r_۲ \int \int P^{n_۰ d-۳}_{mix}(y^{(۳)}, A) P_۲(x, dy^{(۳)}) P_۱(x, dy^{(۱)}) \\
&\vdots \\
&\geq r_۱ r_۲ \dots r_d \int \dots \int P^{(n_۰-۱)d}_{mix}(y^{(n)}, A) P_n(y^{(n-۱)}, dy^{(n)}) \dots P_۱(x, dy^{(۱)}) \\
&= r_۱ r_۲ \dots r_d \int P^{(n_۰-۱)d}_{mix}(y, A) P_{comp}(x, dy)
\end{aligned}$$

اگر $n_۰ > ۱$ باشد، آرگومان بالا $n_۰ - ۱$ مرتبه تکرار می‌شود

$$\begin{aligned}
P^{n_۰ d}_{mix,r}(x, A) &\geq r_۱ r_۲ \dots r_d \int P^{(n_۰-۱)d}_{mix}(y, A) P_{comp}(x, dy) \\
&\geq (r_۱ r_۲ \dots r_d)^۲ \int \int P^{(n_۰-۲)d}_{mix}(z, A) P_{comp}(y, dz) P_{comp}(x, dy) \\
&= (r_۱ r_۲ \dots r_d)^۲ \int P^{(n_۰-۲)d}_{mix}(y, A) P^۲_{comp}(x, dy) \\
&\geq (r_۱ r_۲ \dots r_d)^۳ \int \int P^{(n_۰-۳)d}_{mix}(z, A) P_{comp}(y, dz) P^۲_{comp}(x, dy) \\
&= (r_۱ r_۲ \dots r_d)^۳ \int P^{(n_۰-۳)d}_{mix}(y, A) P^۳_{comp}(x, dy) \\
&\vdots \\
&\geq (r_۱ r_۲ \dots r_d)^{n_۰} P^{n_۰}_{comp}(x, A) \geq (r_۱ r_۲ \dots r_d)^{n_۰} \epsilon Q(A)
\end{aligned}$$

با توجه به رابطه (۱.۳) به این نتیجه می‌رسیم که P_{mix} برای هر احتمال انتخابی، ارگودیک یکنواخت است. □

به‌روز رسانی مولفه به مولفه برای ایجاد ارگودیک یکنواخت P_{mix} ، با توجه به شکل Mtd مشکل است. نتیجه بعدی به‌طور مستقیم از قضایای ۱.۲.۳ و ۲.۲.۳ و مشاهده این‌که P_{RS} و P_{RQ} حالت خاصی از P_{mix} و P_C حالت خاصی از P_{comp} است، حاصل می‌شود. برا دیدن نتایج مرتبط می‌توانید به لاتوس‌زینسکی، رابرتز و روزنتال (۲۰۱۳) و رابرتز و روزنتال (۱۹۹۷) مراجعه کنید.

قضیه ۳.۲.۳. *اگر P_C ارگودیک یکنواخت باشد، P_{RS} و P_{RQ} برای هر احتمال انتخابی ارگودیک یکنواخت هستند.*

برهان. چون P_C حالتی از P_{comp} است و اگر P_C ارگودیک یکنواخت باشد، بنابراین P_{comp} نیز ارگودیک یکنواخت است و طبق قضیه ۲.۲.۳، P_{mix} نیز ارگودیک یکنواخت می‌باشد. از آن‌جایی‌که P_{RS} و P_{RQ} حالتی از P_{mix} هستند، بر اساس قضیه ۱.۲.۳، ارگودیک یکنواخت بودن P_{RS} و P_{RQ}، ثابت می‌شود. □

۱.۲.۳ نمونه‌گیری‌های مستقل مولفه به مولفه

واضح است که بسیاری از زنجیرهای مارکوف مولفه به مولفه ارگودیک یکنواخت نخواهد بود. به عنوان مثال، متروپولیس قدم‌زدن تصادفی در $\mathbb{R}$ ارگودیک یکنواخت نیست، برای اثبات به منگرسن و تویدی (۱۹۹۶) مراجعه کنید. از این‌رو هنگامی‌که، از چنین زنجیرهایی برای ساخت بلوک‌های الگوریتم مولفه به مولفه استفاده می‌شود، انتظار نداریم تا یک زنجیر مارکوف ارگودیک یکنواخت تولید شود. از سوی دیگر منگرسن و تویدی (۱۹۹۶) نشان دادند که نمونه‌های متروپولیس–هستینگز مستقل با بعد کامل، می‌توانند ارگودیک یکنواخت باشند. بنابراین، ما به بررسی الگوریتم مولفه به مولفه‌ای، که به‌روز رسانی هر مولفه، یک الگوریتم متروپولیس–هستینگز با پیشنهادی در حالت مستقل است، می‌پردازیم.

در این صورت، نمونه ترکیب، P_{CIS}، نمونه دنباله تصادفی، P_{RQIS}، و نمونه اسکن تصادفی، P_{RSIS}، همگی آن‌ها نمونه‌های مستقل مولفه به مولفه CWIS، به شمار می‌آیند. ما علاقه‌مند به ایجاد شرایطی هستیم که CWIS ارگودیک یکنواخت شود و با استفاده از قضیه ۳.۲.۳ کافی است که CIS را در نظر بگیریم. توجه داشته باشید که به‌روز رسانی P_{CIS} معمولا گاهی‌اوقات، ترکیبی از چگالی پیشنهادی مولفه به مولفه رد و پذیرش (کسلا و برگر، ۲۰۰۲) خواهد بود. P_{CIS} واقعا یک نمونه مستقل از همه نیست و نتایج منگرسن و تویدی (۱۹۹۶) قابل اجرا نیستند. با این حال با گسترش نگرش منگرسن و تویدی (۱۹۹۶) کار روی MHIS با P_{CIS} راحت خواهد بود.

فرض کنید P_i یک چگالی روی χ باشد که نشان‌دهنده حالت مستقل چگالی پیشنهادی برای i امین به‌روز رسانی $i = ۱, \ldots, d$ باشد. اگر فرض کنیم $p(x) = \prod_{i=۱}^{d} p_i(x_i)$ یک چگالی روی χ و $\epsilon > ۰$، که $p(x) \geq \epsilon\pi(x)$ وجود داشته باشد، آیا یک شرط کافی برای ارگودیک یکنواخت P_{CIS} است؟

اگر تلاش کنیم تا به‌طور مستقیم استدلال منگرسن و تویدی (۱۹۹۶) را تعمیم دهیم با ۲^d مورد برای در نظر گرفتن روبرو می‌شویم. از این‌رو این روش بی‌ثمر است. با این حال، با یک نگرش متفاوت قادر

هستیم تا یک جفت از شرایط، برای ارگودیک یکنواخت CWIS را به‌دست آوریم.

برای توصیف نتایج، نیاز به یک نماد جدید داریم. فرض کنید یک زیرنویس، نشان‌دهنده موقعیت یک مولفه برداری و پرانتز بالانویس، نشان‌دهنده مرحله در یک زنجیر مارکوف باشد. علاوه بر این برای هر $i = ۱, \dots, d$، فرض کنید $x_{[i]} = (x_۱, \dots, x_i)$ و $x^{[i]} = (x_i, \dots, x_d)$ ، و $x_{[۰]}$ و $x^{[d+۱]}$ صفر باشد (برداری با بعد صفر) . نتایج برای CWIS به شرح زیر است.

قضیه ۴.۲.۳. *هسته P_{CIS} با چگالی پیشنهادی p_i را برای $i = ۱, \dots, d$ در نظر بگیرید.*
$p(x) = \prod_{i=۱}^{d} p_i(x_i)$ یک چگالی روی χ تعریف کنید. علاوه بر این فرض کنید $\delta > ۰$، که $p(x) \geq \delta\pi(x)$، برای همه $x \in \chi$ وجود داشته باشد و $\epsilon > ۰$ که برای هر $x, y \in \chi$ با $\pi(x) > ۰$ و $\pi(y) > ۰$ داریم

$$\begin{aligned}\pi(x)\pi(y) &= \pi(x_{[i]}, x^{[i+۱]})\pi(y_{[i]}, y^{[i+۱]}) \\ &\geq \epsilon\pi(x_{[i]}, y^{[i+۱]})\pi(y_{[i]}, x^{[i+۱]}) > ۰\end{aligned} \tag{۴.۳}$$

آن‌گاه برای هر $x \in \chi$ و $A \in \mathcal{B}$،

$$P_{CIS}(x, A) \geq \delta\epsilon^{\lfloor d/۲ \rfloor}\pi(A).$$

بنابراین P_{CIS} ارگودیک یکنواخت است. از این‌رو P_{RQIS} و P_{RSIS} نیز ارگودیک یکنواخت برای هر احتمال انتخابی هستند.

برهان. توجه کنید، در صورتی می‌توان $x = y$ در نظر گرفت که $\epsilon \leq ۱$ باشد. حال برای هر $x, y \in \chi$،

$$\beta(x, y) = \prod_{i=۱}^{d} p_i(y_i)\alpha_i((y_{[i-۱]}, x^{[i]}), y_i)$$

تعریف کنید که، α_i احتمال پذیرش i امین به‌روز رسانی مولفه به مولفه است.

می‌خواهیم نشان دهیم که برای همه $x, y \in \chi$، $\rho > ۰$ به‌طوری‌که $\beta(x, y) \geq \rho\pi(y)$، وجود دارد. همچنین اگر $P(x, y) \geq \beta(x, y)dy$ باشد، زنجیر ارگودیک یکنواخت است.

برای هر x, y می‌توانیم مجموعه شاخص $\{۱, \dots, d\}$ را به $I_۰$ و $I_۱$ تقسیم کنیم و تعریف کنیم

$$I_۰(x, y) = \{i : \alpha_i((y_{[i-۱]}, x^{[i]}), y_i) < ۱\}$$

$$I_۱(x, y) = \{i : \alpha_i((y_{[i-۱]}, x^{[i]}), y_i) = ۱\}$$

آن‌گاه

$$\begin{aligned}\beta(x, y) &= \prod_{i=۱}^{d} p_i(y_i)\alpha_i((y_{[i-۱]}), x^{[i+۱]}), y_i) \\ &= \prod_{i \in I_۱} p_i(y_i) \prod_{i \in I_۰} p_i(x_i)\frac{\pi(y_{[i-۱]}, y_i, x^{[i+۱]})}{\pi(y_{[i-۱]}, x_i, x^{[i+۱]})}.\end{aligned} \tag{۵.۳}$$

این تعریف ساده‌ای است، تا این‌که بیان دیگری برای شاخص مجموعه $I_۰$ و $I_۱$ پیدا کنیم. یک عدد صحیح نامنفی k را به‌صورت زیر تعریف می‌کنیم

- اگر $\alpha_۱ < ۱$ و $\alpha_d < ۱$ باشد آن‌گاه تعداد کل تغییرات، k، بین $\alpha = ۱$ و $\alpha < ۱$ قرار می‌گیرد که $k \leq d - ۱$ است.

- اگر تنها یکی از α_1 و α_d کم‌تر از یک باشند، $k-1$ تغییرات وجود دارد که $k \leq d$ است.
- اگر $\alpha_1 = \alpha_d = 1$ باشد آن‌گاه تعداد تغییرات $k-2$ است که $k \leq d+1$.

توجه داشته باشید که k تنها یک عدد، در هر نمونه است. حال d_1 و d_k را به‌صورت زیر تعریف می‌کنیم

$$d_1 = \begin{cases} d & \text{اگر همه } \alpha_i < 1 \\ min\{i : \alpha_i = 1\} - 1 & \text{در غیر این‌صورت} \end{cases}$$

و

$$d_k = \begin{cases} 0 & \text{اگر همه } \alpha_i < 1 \\ max\{i : \alpha_i = 1\} & \text{در غیر این‌صورت} \end{cases}$$

مجموعه‌ی اعداد صحیح $\{d_0, d_1, d_2, \ldots, d_k, d_{k+1}\}$، به‌طوری‌که

$$0 = d_0 \leq d_1 < d_2 < d_3 < \cdots < d_k \leq d_{k+1} = d$$

تعریف می‌کنیم. در حالت خاص که همه α_i کم‌تر از یک هستند آن‌گاه $k = 0$ و $d_1 = d$، اگر همه α_i ها برابر یک باشند آن‌گاه $k = 2$، $d_1 = 0$ و $d_2 = d$. اگر x و y در χ ثابت، اما دلخواه باشند آن‌گاه ما می‌توانیم I_0 و I_1 را

$$I_0(x,y) = \{d_0 + 1, \ldots, d_1, d_2 + 1, \ldots, d_3, d_4 + 1, \ldots, d_5, \ldots\ldots, d_k + 1, \ldots, d_{k+1}\}$$

$$I_1(x,y) = \{d_1 + 1, \ldots, d_2, d_3 + 1, \ldots, d_4, d_5 + 1, \ldots, d_6, \ldots, d_{k-1} + 1, \ldots, d_k\}$$

بیان کنیم، که نمایش I_0 و I_1 برای نمونه، جایی‌که α_0 و α_d هر دو کم‌تر از یک می‌باشند، باید کاملا واضح باشد. همچنین اگر اجازه دهیم اولین یا آخرین دسته از I_0، صفر باشد، حالت خاص با حالت کلی منطبق می‌شود. نقش صفر برای I_0 به راحتی قابل شناسایی است. اگر $d_1 = 0$ در این‌صورت $\alpha_1 = 1$ است. اولین دسته شاخص‌ها در I_0، $\{d_0 + 1, \ldots, d_1\} = \{1, 0\}$، اگر $d_k = d$ باشد در حالی‌که $\alpha_d = 1$، باید صفر در نظر گرفته شود. آخرین دسته از شاخص‌ها در I_0، $\{d_k + 1, \ldots, d_{k+1}\} = \{d + 1, d\}$ نیز، باید صفر در نظر گرفته شوند.

ما حالا $\rho > 0$ را، به‌طوری‌که برای هر $x, y \in \chi$، $\beta(x,y) \geq \rho\pi(y)$ باشد، پیدا می‌کنیم. ابتدا x و y به‌طوری‌که تمام α_iها کم‌تر از یک هستند، را تصور کنید. آن‌گاه $I_1 = \varnothing$،

$$\beta(x,y) = p(x)\frac{\pi(y)}{\pi(x)} \geq \delta\pi(y)$$

بنابراین باید حالت خاص را در نظر داشته باشیم. فرض کنید که حداقل یکی از $\alpha_i = 1$ باشد، بنابراین ما به یک نماد جدید نیاز خواهیم داشت که در این‌جا معرفی می‌کنیم.

با توجه به $0 = d_0 \leq d_1 < d_2 < d_3 < \cdots < d_k \leq d_{k+1} = d$، I_0 و I_1 را تعریف می‌کنیم. مجموعه

$$s_j = \{d_j + 1, \ldots, d_{j+1}\} \quad \text{برای} \quad j = 0, 1, \ldots, k,$$

به‌طوری‌که $I_1 = \{s_1, s_3, \ldots, s_{k-1}\}$ و $I_0 = \{s_0, s_2, \ldots, s_k\}$. علاوه بر این با توجه به این‌که $z \in X = X_1 \times \cdots \times X_d$ می‌توانیم z را به‌صورت $z = (z_{s_0}, z_{s_1}, \ldots, z_{s_k})$ که $z_{s_j} = (z_{d_j+1}, \ldots, zd_{j+1})$

تقسیم‌بندی کنیم. برای $j = ۰, ۱, \dots, k$، $z_{[j]} = (z_{s_۰}, z_{s_۱}, \dots, z_{s_j})$ و $z^{[j]} = (z_{s_j}, z_{s_{j+۱}}, \dots, z_{s_k})$ را تعریف کنید و فرض کنید $z_{[-۱]}$ و $z^{[k+۱]}$ صفر باشند. توجه کنید

$$p(x_{s_۰}, y_{s_۱}, x_{s_۲}, \dots, y_{s_{k-۱}}, x_{s_k}) = \prod_{i \in I_۱} p_i(y_i) \prod_{i \in I_۰} p_i(x_i).$$

سپس با استفاده از رابطه (۵.۳) و فرض $p(x) \geq \delta\pi(x)$ برای همه x ها

$$\begin{aligned}
\beta(x,y) &= p(x_{s_۰}, y_{s_۱}, x_{s_۲}, y_{s_۳}, x_{s_۴}, \dots, y_{s_{k-۱}}, x_{s_k}) \prod_{j \in \{۰,۲,\dots,k\}} \frac{\pi(y_{[j-۱]}, y_{s_j}, x^{[j+۱]})}{\pi(y_{[j-۱]}, x_{s_j}, x^{[j+۱]})} \\
&\geq \delta\pi(x_{s_۰}, y_{s_۱}, x_{s_۲}, \dots, y_{s_{k-۱}}, x_{s_k}) \prod_{j \in \{۰,۲,\dots,k\}} \frac{\pi(y_{[j-۱]}, y_{s_j}, x^{[j+۱]})}{\pi(y_{[j-۱]}, x_{s_j}, x^{[j+۱]})} \\
&= \delta\pi(x_{s_۰}, y_{s_۱}, x_{s_۲}, \dots, y_{s_{k-۱}}, x_{s_k}) \frac{\pi(y_{s_۰}, x^{[۱]})}{\pi(x_{s_۰}, x^{[۱]})} \prod_{j \in \{۲,۴,\dots,k\}} \frac{\pi(y_{[j-۱]}, y_{s_j}, x^{[j+۱]})}{\pi(y_{[j-۱]}, x_{s_j}, x^{[j+۱]})}.
\end{aligned} \tag{۶.۳}$$

به‌دست می‌آوریم. سپس با یادآوری رابطه (۴.۳) برای هر w و z و همه i ها، جایی‌که $\epsilon > ۰$ وجود داشته باشد داریم

$$\pi(w_{[i]}, w^{[i+۱]})\pi(z_{[i]}, z^{[i+۱]}) \geq \epsilon\pi(w_{[i]}, z^{[i+۱]})\pi(z_{[i]}, w^{[i+۱]}).$$

با شناسایی $w = (x_{s_۰}, y_{s_۱}, x_{s_۲}, \dots, y_{s_{k-۱}}, x_{s_k})$ و $z = (y_{s_۰}, x^{[۱]})$ که کران کم‌تری از RHS با توجه به رابطه (۶.۳) حاصل می‌شود

$$\delta\epsilon\pi(y_{[۱]}, x_{s_۲}, y_{s_۳}, x_{s_۴}, \dots, y_{s_{k-۱}}, x_{s_k}) \prod_{j \in \{۲,۴,\dots,k\}} \frac{\pi(y_{[j-۱]}, y_{s_j}, x^{[j+۱]})}{\pi(y_{[j-۱]}, x_{s_j}, x^{[j+۱]})}.$$

در ادامه به‌دست می‌آید

$$\begin{aligned}
&\delta\pi(x_{s_۰}, y_{s_۱}, x_{s_۲}, \dots, y_{s_{k-۱}}, x_{s_k}) \prod_{j \in \{۲,۴,\dots,k\}} \frac{\pi(y_{[j-۱]}, y_{s_j}, x^{[j+۱]})}{\pi(y_{[j-۱]}, x_{s_j}, x^{[j+۱]})} \\
&\geq \delta\epsilon\pi(y_{[۱]}, x_{s_۲}, y_{s_۳}, x_{s_۴}, \dots, y_{s_{k-۱}}, x_{s_k}) \prod_{j \in \{۲,۴,\dots,k\}} \frac{\pi(y_{[j-۱]}, y_{s_j}, x^{[j+۱]})}{\pi(y_{[j-۱]}, x_{s_j}, x^{[j+۱]})} \\
&\geq \delta\epsilon^۲\pi(y_{[۳]}, x_{s_۴}, y_{s_۵}, x_{s_۶}, \dots, y_{s_{k-۱}}, x_{s_k}) \prod_{j \in \{۴,۶,\dots,k\}} \frac{\pi(y_{[j-۱]}, y_{s_j}, x^{[j+۱]})}{\pi(y_{[j-۱]}, x_{s_j}, x^{[j+۱]})} \\
&\vdots \\
&\geq \delta\epsilon^{k/۲}\pi(y_{[k-۱]}, x_{s_k}) \frac{\pi(y_{[k-۱]}, y_{s_k})}{\pi(y_{[k-۱]}, x_{s_k})} \\
&= \delta\epsilon^{k/۲}\pi(y)
\end{aligned}$$

به خاطر بیاورید که $k \leq d+۱$ است اما توجه داشته باشید که برابری به این معناست که $k-۲$ تغییرات وجود داشته است که حالت اول دسته $I_۰$، صفر است و ما تنها به کم‌ترین ϵ برای سمت راست نامساوی بالا نیازمندیم. بنابراین $\beta(x,y) \geq \rho\pi(y)$ با $\rho = \delta\epsilon^{\lfloor d/۲ \rfloor}$ برقرار است، و نتیجه می‌گیریم که زنجیر ارگودیک یکنواخت است. □

دو نتیجه‌گیری زیر نشان‌دهنده این است که شرایط قضیه به راحتی پذیرفته شده است.

نتیجه ۵.۲.۳. *هسته P_{CIS} با چگالی پیشنهادی p_i برای $i = ۱, \ldots, d$ در نظر بگیرید. $p(x) = \prod_{i=۱}^{d} p_i(x_i)$ یک چگالی روی χ تعریف کنید. علاوه بر این فرض کنید برای همه $x \in \chi$، $\delta > ۰$، وجود دارد که $p(x) \geq \delta\pi(x)$. همچنین اگر جفت توابع مثبت g_i و h_i روی X_i، برای $i = ۱, \ldots, d$ برای هر $x \in \chi$ در نظر بگیرید به‌طوری‌که*

$$\prod_{i=۱}^{d} g_i(x_i) \leq \pi(x) \leq \prod_{i=۱}^{d} h_i(x_i) \tag{۷.۳}$$

و $۰ < inf_{x_i \in X_i} \{\frac{g_i(x_i)}{h_i(x_i)}\}$ به ازای هر $i = ۱, \ldots, d$، آنگاه P_{CIS} و P_{RQIS} و P_{RSIS} همگی ارگودیک یکنواخت هستند.

برهان. برای اثبات این نتیجه باید نشان دهیم که رابطه (۷.۳) در واقع، به همان رابطه (۴.۳) اشاره می‌کند. فرض کنید برای $i = ۱, \ldots, d$، $\rho_i = inf_{x_i \in X_i} \{\frac{g_i(x_i)}{h_i(x_i)}\}$، به‌طوری‌که $\rho_i > ۰$. آنگاه برای هر $x, y \in \chi$ به ازای $i = ۱, \ldots, d-۱$ داریم

$$\frac{\pi(x_{[i]}, x^{[i+۱]})\pi(y_{[i]}, y^{[i+۱]})}{\pi(x_{[i]}, y^{[i+۱]})\pi(y_{[i]}, x^{[i+۱]})} \geq \prod_{j=۱}^{i} \frac{g_j(x_j)g_j(y_j)}{h_j(x_j)h_j(y_j)} \prod_{j=i+۱}^{d} \frac{g_j(x_j)g_j(y_j)}{h_j(y_j)h_j(x_j)} \geq \prod_{j=۱}^{d} \rho_j^۲$$

بنابراین رابطه (۴.۳) با $\epsilon = (\rho_۱ \ldots \rho_d)^۲$ برقرار است. □

شرایط نتیجه ۵.۲.۳ به شکل ضعیفی از وابستگی در توزیع هدف نیاز دارد. واضح‌ترین حالت خاص زمانی است که مولفه‌های π تواما مستقل باشند که در رابطه (۷.۳)، برابری در هر دو طرف برقرار است.

نتیجه ۶.۲.۳. *هسته P_{CIS} با چگالی پیشنهادی p_i برای $i = ۱, \ldots, d$ در نظر بگیرید. $p(x) = \prod_{i=۱}^{d} p_i(x_i)$ یک چگالی روی χ تعریف کنید. اگر $۰ < a \leq b < \infty$ و $c > ۰$ وجود داشته باشد به‌طوری‌که $a \leq \pi(x) \leq b$ و $p(x) \geq c$، برای ϖهای همه مقادیر x برقرار باشد، آنگاه P_{CIS} و P_{RQIS} و P_{RSIS} همگی ارگودیک یکنواخت هستند.*

برهان. کلیه شرایط قضیه ۴.۲.۳ و رابطه ۶.۳ با در نظر گرفتن $\delta = c/b$ و $\epsilon = (a/b)^۲$، برقرا است. آنگاه ارگودیک یکنواخت بودن P_{CIS} و P_{RQIS} و P_{RSIS} حاصل می‌شود. □

۲.۲.۳ ماکسیمم درستنمایی برای مدل‌های آمیخته

فرض کنید $Y_i = \{Y_{i۱}, \ldots, Y_{im_i}\}$ نشان‌دهنده یک برداری از داده‌های قابل مشاهده و U_i نشان‌دهنده i امین اثر تصادفی غیرقابل مشاهده برای $i = ۱, \ldots, k$ باشد.

فرض کنید $U = (U_۱, \ldots, U_k)$ و Y_i مستقل با توزیع شرطی مشخص شده روی $U = u$ در نظر بگیرید، به‌طوری‌که چگالی توام $Y = \{Y_{ij} : j = ۱, \ldots, m_i\,;\, i = ۱, \ldots, k\}$ به‌صورت

$$f(y|u;\theta_۱) = \prod_{i=۱}^{k} \prod_{j=۱}^{m_i} f(y_{ij}|u_i;\theta_۱).$$

θ_1 نشان‌دهنده یک برداری از پارامترهاست. U_i نیز مستقل است و توزیع آن‌ها لزوما نرمال نمی‌باشد. بنابراین چگالی توام مولفه‌های U به‌صورت، $h(u;\theta_2) = \prod_{i=1}^{k} h(u_i;\theta_2)$ تعریف می‌شود و، درستنمایی

$$L(\theta|y) = \int f(y|u;\theta_1)h(u;\theta_2)\,du.$$

اغلب تحلیل این انتگرال رام‌نشدنی است، به‌طوری‌که محاسبه برآورد ماکسیمم درستنمایی و استاندارد خطا می‌تواند چالش برانگیز باشد. با این حال چندین الگوریتم مبتنی بر مونت‌کارلو، مانند مونت‌کارلوی نیوتن رافسون، مونت‌کارلوی ماکسیمم درستنمایی و مونت‌کارلوی EM [1] وجود دارد، که برای پیدا کردن برآورد ماکسیمم درستنمایی پارامتر مجهول، $\theta = (\theta_1, \theta_2)$ مناسب می‌باشند (سافو، جیک و جونز، ۲۰۰۵؛ هابرت، ۲۰۰۰ و مک‌کالاک، ۱۹۹۷).

از ویژگی‌های مشترک هر سه الگوریتم این است که نیاز به شبیه‌سازی از توزیع هدف دارند. یعنی توزیع شرطی اثرات تصادفی، با توجه به داده به‌دست می‌آید به‌طوری‌که برای مقدار معینی از θ

$$h(u|y;\theta) \propto f(y|u;\theta_1)h(u;\theta_2)$$

چهار زنجیر مارکوف، با داشتن $h(u|y;\theta)$ به عنوان چگالی مانا که شامل سه نمونه مولفه به مولفه مستقل، CIS، RQIS و RSIS با چگالی پیشنهادی $h(u_i;\theta_2)$ برای $i = 1,\ldots,k$، و یک نمونه متروپولیس-هستینگز مستقل با بعد کامل، (MHIS) با چگالی پیشنهادی انتخاب شده، از توزیع حاشیه‌ایی $h(u;\theta_2)$ در نظر می‌گیریم. برای این منظور نتیجه به شرح زیر است.

قضیه ۷.۲.۳. *اگر $B(y;\theta_1) < \infty$ وجود داشته باشد به‌طوری‌که برای تمام u، $f(y|u;\theta_1) \le B(y;\theta_1)$. آن‌گاه چهار زنجیر مارکوف توصیف شده در بالا MHIS، CIS، RQIS و RSIS با داشتن $h(u|y,\theta)$ به عنوان چگالی مانا، ارگودیک یکنواخت هستند.*

برهان. MHIS را با داشتن چگالی پیشنهادی $h(u;\theta_2)$ در نظر بگیرید. احتمال پذیرش برای یک چگالی پیشنهادی، پرش از u به u^* به‌صورت

$$\alpha(u,u^*) = \min\left\{1, \frac{f(y|u^*;\theta_1)}{f(y|u;\theta_1)}\right\}$$

ساده می‌شود. این زنجیر مارکوف به‌وسیله قضیه ۲.۱ منگرسن و توییدی (۱۹۹۶) ارگودیک یکنواخت است که نسبت چگالی پیشنهادی به هدف آن

$$\frac{C(y;\theta)h(u;\theta_2)}{f(y|u;\theta_1)h(u;\theta_2)} = \frac{C(y;\theta)}{f(y|u;\theta_1)} \ge \frac{C(y;\theta)}{B(y;\theta)}$$

تعریف می‌شود. در این رابطه $C(y;\theta) = \int f(y|u;\theta_1)h(u;\theta_2)du$.

هر سه نمونه‌گیری مستقل مولفه به مولفه CIS، RQIS، RSIS با چگالی پیشنهادی $h(u_i;\theta_2)$ برای $i = 1,\ldots,k$، به‌وسیله نتیجه ۵.۲.۳ و رابطه ۴.۳

$$h(u|y;\theta) \propto f(y|u;\theta_1)h(u;\theta_2) = \prod_{i=1}^{k} h(u_i;\theta_2)\prod_{j=1}^{n_i} f(y_{ij}|u_i;\theta_1)$$

ارگودیک یکنواخت بودن هر سه نمونه‌گیری اثبات می‌شود. □

[1] MC expectation maximization

ما عملکرد تجربی CWIS، MHIS و ارگودیک هندسی یک نمونه متروپولیس قدم‌زدن تصادفی با بعد کامل را، در یک مثال مرتبط، در فصل بعد مقایسه می‌کنیم.

۳.۳ ارگودیک هندسی تحت آمیختگی و ترکیب

P_{mix} را در نظر بگیرید و فرض کنید که هر یک از هسته‌های P_i ارگودیک هندسی باشند، تابع نامنفی M_i و $t_i \in (۰, ۱)$ وجود دارد به‌طوری‌که

$$\|P_i(x, .) - \varpi(.)\| \leq M_i(x) t_i^n$$

آن‌گاه نامساوی مثلث اشاره دارد که

$$\begin{aligned}
\|P_{mix}(x, .) - \varpi(.)\| &= \|r_۱ P_۱ + \cdots + r_d P_d - \varpi(.)\| \\
&= \|r_۱(P_۱ - \varpi(.)) + r_۲(P_۲ - \varpi(.)) + \cdots + r_d(P_d - \varpi(.))\| \\
&\leq \sum_{i=۱}^{d} r_i \|P_i - \varpi(.)\| \leq \sum_{i=۱}^{d} r_i M_i(x) t_i^n \\
&\leq max\{t_i\}^n \sum_{i=۱}^{d} r_i M_i(x) \\
&\leq \sum_{i=۱}^{d} r_i M_i(x) \left[Max\{t_i\}\right]^n
\end{aligned}$$

بنابراین، اگر هر P_i ارگودیک هندسی باشد، P_{mix} نیز ارگودیک هندسی است. این، تفاوت بین ایجاد ارگودیک هندسی و یکنواخت را نشان می‌دهد. به خاطر بیاورید که اگر یکی از P_i ها ارگودیک یکنواخت باشد، ارگودیک یکنواخت بودن P_{mix} را نتیجه می‌دهد. همچنین بلافاصله این نتیجه حاصل می‌شود که P_{RQ} ارگودیک هندسی است اگر هر یک از نمونه‌های ترکیب ارگویک هندسی باشند.

این نتایج برای P_{RS} صادق نیست. در این حالت هر یک از P_iها به‌طور معمول حتی، ϖ-تحویل‌ناپذیر نیستند، زیرا که بر اساس تعریف P_{RS}، یک مولفه را با یک احتمالی انتخاب می‌کردیم و بقیه مولفه‌ها را ثابت در نظر می‌گرفتیم، این به این معنی است که احتمال جا به جایی از یک مولفه به مولفه دیگر وجود ندارد.

قضیه ۱.۳.۳. *فرض کنید P_{mix} با توجه به ϖ برای همه احتمال انتخاب $r \in \mathbb{P}^d$ برگشت‌پذیر باشد. اگر P_{mix} ارگودیک هندسی برای برخی احتمال انتخاب باشد آن‌گاه آن ارگودیک هندسی برای همه احتمال انتخاب است.*

برهان. فرض کنید $P_{mix,r}$ نشان‌دهنده P_{mix} برای انتخاب خاصی از $r \in \mathbb{P}^d$ باشد، همچنین فرض کنید $r^* \in \mathbb{P}^d$, r و این‌که P_{mix,r^*} ارگودیک هندسی است. آن‌گاه اگر $A \in \mathcal{B}$

$$P_{mix,r}(x, A) \geq \min\left\{\frac{r_۱}{r_۱^*}, \frac{r_۲}{r_۲^*}, \ldots, \frac{r_d}{r_d^*}\right\} P_{mix,r^*}(x, A)$$

بر اساس قضیه ۱ از جونز، رابرتز و روزنتال (۲۰۱۳) این اثبات کامل می‌شود. □

توجه داشته باشید که حالت خاصی از قضیه ۱.۳.۳ این است که اگر P_{RS} ارگودیک هندسی برای برخی احتمال انتخاب باشد آنگاه آن ارگودیک هندسی برای همه احتمالات انتخاب است. چون کلیه شرایط قضیه ۱.۳.۳ برای P_{RS} صدق می‌کند. برای نتایج مرتبط جونز، رابرتز و روزنتال (۲۰۱۴) را ببینید. همچنین قضیه ۱.۳.۳ برای نمونه‌های دنباله اسکن تصادفی کاربرد ندارد، چونکه همه احتمالات انتخاب برگشت‌پذیر نیستند.

۱.۳.۳ تنظیمات دو متغیره

ما حالتی که ϖ با توجه به $\mu_1 \times \mu_2$ دارای چگالی $\pi(x,y)$ است و از $X_1 \times X_2 \subseteq \mathbb{R}^{b_1} \times \mathbb{R}^{b_2}$ پشتیبانی می‌کند، را در نظر می‌گیریم. فرض کنید $\pi_{X|Y}(x|y)$ و $\pi_{Y|X}(y|x)$ چگالی‌های شرطی کامل و π_X و π_Y چگالی‌های حاشیه‌ایی حاصل از π باشند (ϖ_X و ϖ_Y توزیع‌های حاشیه‌ای هستند). این تنظیمات، هر چند به‌طور کلی کم‌تر از بخش قبلی متداول است اما، دارای برنامه‌های کاربردی بسیار عملی می‌باشد. به عنوان مثال، آن، شالوده‌ایی برای روش‌های تقویت داده (هابرت، ۲۰۱۱؛ تانر و وونگ، ۱۹۸۷)، بسیاری از روش‌های MCMC به‌خصوص مدل‌های آماری مرتبط، است (رومن و هابرت، ۲۰۱۲؛ روی و هابرت، ۲۰۰۷ و جانسون و جونز، ۲۰۱۰).

در این جا دو تنظیمات را می‌خواهیم در نظر بگیریم:

۱) با حالتی که نمونه‌گیری از $\pi_{X|Y}$ و $\pi_{Y|X}$ امکان‌پذیر باشد، شروع می‌کنیم.

۲) حالتی که یکی از به‌روز رسانی گیبز، به‌وسیله یک به‌روز رسانی متروپولیس–هستینگز جایگزین می‌شود.

زمانی‌که نمونه‌گیری از $\pi_{X|Y}$ و $\pi_{Y|X}$ ساده باشد هسته مارکوف تشکیل شده از ترکیب، P_{GS} نامیده می‌شود که معمولا نمونه‌گیر گیبز، GS دارای (Mtd)

$$h_{GS}(x\prime, y\prime|x,y) = \pi_{X|Y}(x\prime|y)\pi_{Y|X}(y\prime|x\prime) \qquad (۸.۳)$$

است. در این حالت ابتدا x و سپس y به‌روز می‌شود البته شیوه دیگر به‌روز رسانی که همچنین یک نمونه‌گیر گیبز است را با $\tilde{P}_{GS}$ نشان می‌دهند، که ترتیب به‌روز رسانی x و y، تغییر می‌کند. از طرفی هر یک از دنباله‌های حاشیه‌ایی $\{X^{(n)}\}$ و $\{Y^{(n)}\}$ به ترتیب، یک مرحله هسته مارکوف P_X و P_Y با $Mtds$

$$h_X(x\prime|x) = \int \pi_{X|Y}(x\prime|y)\pi_{Y|X}(y|x)\mu_2(dy)$$

و

$$h_Y(y\prime|y) = \int \pi_{Y|X}(y\prime|x)\pi_{X|Y}(x|y)\mu_1(dx) \qquad (۹.۳)$$

دارا هستند. علاوه بر این ساده است تا ببینید که P_Y^m یک Mtd، m مرحله‌ای $h_Y^m(y\prime|y)$، به عنوان انجام P_X^m، P_{GS}^m و $\tilde{P}_{GS}^m$ می‌پذیرد. توجه داشته باشید که π_X برای $\{X^{(n)}\}$ و π_Y برای $\{Y^{(n)}\}$ ماناست. به خوبی معلوم است که P_X، P_Y، P_{GS} و $\tilde{P}_{GS}$ تماما با همان نرخ کیفی همگرا هستند. برای بررسی جزییات به دایاکنس،خیر و سالوف–کاست (۲۰۰۸)، رابرت (۱۹۹۵) و رابرتز و روزنتال

(۲۰۰۱) مراجعه کنید. این روابط، به‌طور معمول در تجزیه و تحلیل نمونه‌گیری‌های گیبز و مدل‌های آماری مربوطه، که در آن اغلب تحلیل P_Y یا P_X از P_{GS} به‌دلیل کم بودن بعد، ساده است، مورد استفاده قرار می‌گیرد.

با قردادن این مشاهدات می‌توان چنین استنتاج کرد که اگر یکی از هسته‌های P_X و P_Y و P_{GS} یا $\tilde{P}_{GS}$ ارگودیک هندسی باشد آنگاه بقیه و همچنین نمونه دنباله تصادفی گیبز، P_{RQGS} نیز، ارگودیک هندسی هستند.

اولین نتیجه این زیر بخش به نرخ همگرایی P_X، P_Y، P_{GS} و $\tilde{P}_{GS}$ برای نمونه‌گیری اسکن تصادفی گیبز، P_{RSGS} مربوط می‌شود. توجه کنید که به‌وسیله رابطه (۲.۳) و (۹.۳) داریم که

$$P_Y\, W(y) = \int_{X_2} W(y\prime) h_Y(y\prime|y) \mu_2(dy\prime)$$

قضیه ۲.۳.۳. *فرض کنید $\lambda < 1$، $W : X_2 \longrightarrow \mathbb{R}^+$ و $b < \infty$ وجود داشته باشد به‌طوری‌که*

$$P_Y\, W(y) \leq \lambda W(y) + b \qquad (۱۰.۳)$$

فرض کنید $C_d = \{y : W(y) \leq d\}$، $g : X_2 \longrightarrow \mathbb{R}^+$ و یک $d_\circ > \circ$ وجود داشته باشد، به‌طوری‌که برای برخی $m \geq 1$

$$h_Y^m(y\prime|y) \geq g(y\prime) \qquad \text{برای همه} \quad y \in C_d \text{ و } d \geq d_\circ. \qquad (۱۱.۳)$$

آنگاه P_Y و P_X و P_{GS} و $\tilde{P}_{GS}$ به عنوان P_{RQGS} و P_{RSGS} ارگودیک هندسی هستند.

برای اثبات قضیه ۲.۳.۳ نیاز به ذکر دو لم اولیه است که ابتدا آن‌ها را بیان می‌کنیم و سپس به اثبات قضیه می‌پردازیم.

لم ۳.۳.۳. *فرض کنید W یک تابع حقیقی مقدار روی χ و $\circ < \lambda < 1$، $b < \infty$ ثابت باشند که*

$$P_X W(x) \leq \lambda W(x) + b \qquad \text{برای همه} \quad x \in \chi \qquad (۱۲.۳)$$

۱- اگربرای همه $x \in \chi$ ، $W(x) \geq 1$ باشد، مجموعه $V(x) = W(x)$، $\gamma = (\lambda + 1)/2$ و $k = b$.

۲- اگر برای همه $x \in \chi$، $W(x) \geq \circ$ باشد، مجموعه $V(x) = 1 + W(x)$، $\gamma = (\lambda + 1)/2$ و $k = b + (1 - \lambda)$. فرض کنید $C = \{x : V(x) \leq k/(1 - \gamma)\}$ آنگاه شرط رانش (۳.۳) برقرار است.

برهان. توجه کنید که برای هر $\circ < \lambda < 1$، $k < \infty$ و تابع مثبت V اگر $\gamma = (\lambda+1)/2$ باشد، آنگاه

$$\lambda V(x) + k = (2\gamma - 1)V(x) + k = \gamma V(x) - (1 - \gamma)V(x) + k.$$

با فرض $C = \{x : V(x) \leq k/(1 - \gamma)\}$. اگر $x \notin C$، آنگاه $V(x) > k/(1 - \gamma)$ و از این‌رو

$$\lambda V(x) + k \leq \gamma V(x) - (1 - \gamma)\frac{k}{1 - \gamma} + k = \gamma V(x)$$

اما برای همه $x \in \chi$، $\lambda V(x) + k \leq \gamma V(x) + k$. بنابراین $\lambda V(x) + k \leq \gamma V(x) + k I_C(x)$ است. ما برای کامل کردن اثبات نیاز داریم تا نشان دهیم که فرض (۱۲.۳)، اگر $W(x) \geq 1$، $V(x) = W(x)$

و $k = b$ باشد، دلالت بر $PV(x) \leq \lambda V(x) + k$ دارد. از طرفی دیگر فرض کنید $W(x) \geq ۰$، مجموعه $V(x) = ۱ + W(x)$ و $k = b + (۱ - \lambda)$ باشد. آن‌گاه

$$PV(x) = P_X W(x) + ۱ \leq \lambda W(x) + ۱ + b = \lambda V(x) - \lambda + ۱ + b = \lambda V(x) + k$$

□

هر سه زنجیر مارکف و همچنین m امین مرحله Mtd را که با h^m_{GS}، h^m_X و h^m_Y نشان می‌دهند را به خاطر بیاورید. ما به لم مقدماتی دیگری قبل از اثبات قضیه ۲.۳.۳ نیاز داریم.

لم ۴.۳.۳. *اگر $m \geq ۲$ باشد، آن‌گاه*

$$h^m_{GS}(x', y'|x, y) = \pi_{Y|X}(y'|x') \int_{x_۲} \pi_{X|Y}(x'|z) h^{m-۱}_Y(z|y) \mu_۲(dz).$$

برهان. فرض کنید $m = ۲$ باشد آن‌گاه

$$\begin{aligned}
h^۲_{GS}(x', y'|x, y) &= \int_{X_۱} \int_{X_۲} \pi_{X|Y}(x'|\upsilon) \pi_{Y|X}(y'|x') \pi_{X|Y}(u|y)_{Y|X}(\upsilon|u) \mu_۲(d\upsilon) \mu_۲(du) \\
&= \pi_{Y|X}(y'|x') \int_{X_۲} \pi_{X|Y}(x'|\upsilon) \int_{X_۱} \pi_{Y|X}(\upsilon|u) \pi_{X|Y}(u|y) \mu_۱(du) \mu_۲(d\upsilon) \\
&= \pi_{Y|X}(y'|x') \int_{X_۲} \pi_{X|Y}(x'|\upsilon) h_Y(\upsilon|y) \mu_۲(d\upsilon)
\end{aligned}$$

با یک محاسبه مشابه نتیجه برای $m = ۳$ برقرار است و یک دلیل برقراری کامل شدن اثبات است.

$$\begin{aligned}
h^{m+۱}_{GS}(x', y'|x, y) &= \int_{X_۱} \int_{X_۲} h(x', y'|u, \upsilon) h^m(u, \upsilon|x, y) \mu_۲(d\upsilon) \mu_۱(du) \\
&= \int_{X_۱} \int_{X_۲} \pi_{X|Y}(x', \upsilon) \pi_{Y|X}(y'|x') h^m(u, \upsilon|x, y) \mu_۲(d\upsilon) \mu_۱(du) \\
&= \pi_{Y|X}(y'|x') \left(\int_{X_۱} \int_{X_۲} \pi_{X|Y}(x'|\upsilon) \pi_{Y|X}(\upsilon|u) \right. \\
&\qquad \left. \int_{X_۲} \pi_{X|Y}(u|z) h^{m-۱}_Y(z|y) \mu_۲(dz) \mu_۲(d\upsilon) \mu_۱(du) \right) \\
&= \pi_{Y|X}(y'|x') \left(\int_{X_۲} \pi_{X|Y}(x'|\upsilon) \right. \\
&\qquad \left. \int_{X_۲} \int_{X_۱} \pi_{Y|X}(\upsilon|u) \pi_{X|Y}(u|z) h^{m-۱}_Y(z|y) \mu_۲(dz) \mu_۱(du) \mu_۲(d\upsilon) \right) \\
&= \pi_{Y|X}(y'|x') \left(\int_{X_۲} \pi_{X|Y}(x'|\upsilon) \int_{X_۲} h^{m-۱}_Y(z|y) \right. \\
&\qquad \left. \int_{X_۱} \pi_{Y|X}(\upsilon|u) \pi_{X|Y}(u|z) \mu_۱(du) \mu_۲(dz) \mu_۲(d\upsilon) \right) \\
&= \pi_{Y|X}(y'|x') \left(\int_{X_۲} \pi_{X|Y}(x'|\upsilon) \int_{X_۲} h_Y(\upsilon|z) h^{m-۱}_Y(z|y) \mu_۲(dz) \mu_۲(d\upsilon) \right) \\
&= \pi_{Y|X}(y'|x') \left(\int_{X_۲} \pi_{X|Y}(x'|\upsilon) h^m_Y(\upsilon|y) \mu_۲(d\upsilon) \right)
\end{aligned}$$

□

حال با کمک این دو لم قضیه ۲.۳.۳ اثبات می‌شود.

برهان. فرض کنید $V(y) = ۱ + W(y)$ و مجموعه $\gamma = (\lambda + ۱)/۲$ و $k = b + ۱ - \lambda < \infty$ باشد. از لم ۳.۳.۳ به این نتیجه رسیدیم که فرض‌های ما اشاره به شرط رانش زیر دارد

$$P_Y V(y) \leq \gamma V(y) + k I_C(y)$$

که $C = \{y : V(y) \leq k/(۱ - \gamma)\}$. برای این‌که نشان دهیم P_Y ارگودیک هندسی است تنها نیاز به برقراری شرط کوچک‌سازی روی C داریم. در حال حاضر $C = \{y : W(y) \leq k/(۱ - \gamma) - ۱\}$ و $k/(۱ - \gamma) - ۱ = (۲b - \lambda + ۱)/(۱ - \gamma) > ۰$ اشاره بر این دارند که C ناتهی است. حال برای d به اندازه کافی بزرگ، $C \subseteq C_d$ و رابطه (۱۱.۳)، بلافاصله شرط کوچک‌سازی برای $A \subseteq X_۲$ با

$$\epsilon = \int_{X_۲} g(z)\mu_۲(dz) \qquad \text{و} \qquad Q(A) = \epsilon^{-۱} \int_A g(z)\mu_۲(dz)$$

به‌کار برده می‌شود. بنابراین P_Y ارگودیک هندسی است و همچنین P_X و P_{GS} و $\tilde{P}_{GS}$ نیز ارگودیک هندسی هستند.

نمونه‌گیری اسکن تصادفی گیبز را در نظر بگیرید که در هر تکرار، X با احتمال p و Y با احتمال $۱ - p$ به‌روز رسانی می‌شوند. فرض کنید $G(x) = \int_{X_۲} W(y)\pi_{Y|X}(y|x)\mu_۲(dy)$. فرض کنید υ

$$\frac{۱ - p}{p} < \upsilon < \frac{۱ - p}{p^{\lambda}} \tag{۱۳.۳}$$

به‌علاوه $V_۱(x, y) = \upsilon G(x) + W(y)$. آن‌گاه

$$\begin{aligned}
P_{RSGS}V_۱(x,y) &= \int pV_۱(x',y')\pi_{X|Y}(x'|y)\delta(y'-y)\mu_۱(dx')\mu_۲(dy') \\
&\quad + \int (۱-p)V_۱(x',y')\pi_{Y|X}(y'|x)\delta(x'-x)\mu_۱(dx')\mu_۲(dy') \\
&= \int pV_۱(x',y)\pi_{X|Y}(x'|y)\mu_۱(dx') \\
&\quad + \int (۱-p)V_۱(x,y')\pi_{Y|X}(y'|x)\mu_۲(dy') \\
&= p\int [\upsilon(G(x')+W(y)]\pi_{X|Y}(x'|y)\mu_۱(dx') \\
&\quad + (۱-p)\int [\upsilon G(x)+W(y')]\pi_{Y|X}(y'|x)\mu_۲(dy') \\
&= p\upsilon \int G(x')\pi_{X|Y}(x'|y)\mu_۱(dx') \\
&\quad + (۱-p)\int W(y')\pi_{Y|X}(y'|x)\mu_۲(dy') + pW(y) + (۱-p)\upsilon G(x) \\
&= p\upsilon \int G(x')\pi_{X|Y}(x'|y)\mu_۱(dx') + (۱-p)(۱+\upsilon)G(x) + pW(y)
\end{aligned}$$

حال انتگرال باقیمانده را به‌طور جداگانه در نظر بگیرید

$$\begin{aligned}\int \pi_{X|Y}(x'|y)G(x')\mu_1(dx') &= \int \pi_{X|Y}(x'|y)\int W(y')\pi_{Y|X}(y'|x')\mu_2(dy')\mu_1(dx')\\ &= \int\int W(y')\pi_{X|Y}(x'|y)\pi_{Y|X}(y'|x')\mu_2(dy')\mu_1(dx')\\ &= \int W(y')\int \pi_{Y|X}(y'|x')\pi_{X|Y}(x'|y)\mu_1(dx')\mu_2(dy')\\ &= \int W(y')h_Y(y'|y)\mu_2(dy')\\ &\le \lambda W(y)+b\end{aligned}$$

به موقعیت قبلی در بالا برگردید، و در نتیجه

$$P_{RSGS}V_1(x,y) \le p(\upsilon\lambda+1)W(y)+(1-p)(1+\upsilon)G(x)+p\upsilon k.$$

به‌وسیله رابطه (۱۳.۳)، $\beta < 1$ وجود دارد به‌طوری‌که

$$max\left\{p(\upsilon\lambda+1), \frac{(1-p)(1+\upsilon)}{\upsilon}\right\} \le \beta$$

آن‌گاه

$$P_{RSGS}V_1(x,y) \le \beta[W(y)+\upsilon G(x)]+p\upsilon k = \beta V_1(x,y)+p\upsilon b. \qquad (۱۴.۳)$$

فرض کنید مجموعه $H_1 = 1+V_1$، $\gamma_1 = (\beta+1)/2$ و $C = \{(x,y) : H_1(x,y) \le p\upsilon b/(1-\gamma)\}$.
آن‌گاه از لم ۳.۳.۳ و رابطه (۱۴.۳) پیروی می‌کند که شرط کوچک‌سازی برای همه (x,y) برقرار است.

$$P_{RSGS}H_1(x,y) \le \gamma_1 H_1(x,y)+p\upsilon b I_c(x,y)$$

در نهایت باید بیان کنیم که C کوچک است. توجه کنید که

$$P^{2m}_{RSGS}((x,y),\cdot) \ge [p(1-p)]^m P^m_{GS}((x,y),\cdot)$$

و از این‌رو کافی است تا یک شرط کوچک‌سازی برای P^m_{GS} که $m \ge 1$، ایجاد کنیم.
برای $d > 1$، فرض کنید $C_d = \{y : W(y) \le d\}$. کافی است ایجاد کنیم که $X_1 \times C_d$ کوچک است، زمانی‌که $C \subseteq X_1 \times C_d$، به‌طوری‌که اگر $d \ge p\upsilon k/(1-\gamma_1)$. با فرض این‌که یک $d \ge p\upsilon k/(1-\gamma_1)$ و یک تابع g برای برخی $m \ge 1$ و d به اندازه کافی بزرگ وجود داشته باشد، داریم

$$h^m_Y(y'|y) \ge g(y') \qquad \text{برای همه} \qquad y \in C_d.$$

توجه کنید، زمانی‌که رابطه (۱۱.۳) با $m = 1$ برقرار است، ما تنها باید برای حالت $m \ge 2$ در نظر بگیریم. سپس

$$h^2(y\prime|y) = \int_{X_2} h_Y(y\prime|z)h_Y(z|y)\mu_2(dz) \ge \int_{X_2} h_Y(y\prime|z)g(z)\mu_2(dz) := g^*(y\prime)$$

از این‌رو رابطه (۱۱.۳) با $m = ۲$ برقرار است. فرض کنید $m \geq ۲$ باشد و به‌وسیله لم ۴.۳.۳ و رابطه (۱۱.۳) اگر $y \in C_d$

$$\begin{aligned} h^m_{GS}(x', y'|x, y) &= \pi_{Y|X}(y'|x') \int_{X_۲} \pi_{X|Y}(x'|z) h^{m-۱}_Y(z|y) \mu_۲(dz) \\ &\geq \pi_{Y|X}(y'|x') \int_{X_۲} \pi_{X|Y}(x'|z) g(z) \mu_۲(dz) \\ &= \epsilon q(x', y') \end{aligned}$$

با

$$\begin{aligned} \epsilon &= \int\int \pi_{Y|X}(y'|x') \int_{X_۲} \pi_{X|Y}(x'|z) g(z) \mu_۲(dz) \mu_۱(dx') \mu_۲(dy') \\ q(x', y') &= \epsilon^{-۱} \pi_{Y|X}(y'|x') \int_{X_۲} \pi_{X|Y}(x'|z) g(z) \mu_۲(dz) \end{aligned}$$

بنابراین ثابت می‌شود که $X_۱ \times C_d$ برای P^m_{GS} و همچنین $P^{۲m}_{RSGS}$ کوچک است. برهان ارگودیک یکنواخت بودن RSGS کامل می‌شود. □

نتیجه ۵.۳.۳. *یک شرط کافی و ساده برای رابطه (۱۱.۳) وجود دارد؛ فرض کنید یک* $\ell : X_۲ \longrightarrow \mathbb{R}^+$ *که* $\pi_{X|Y}(x|y) \geq \ell(x)$*، برای همه* $(x, y) \in X_۱ \times C_d$ *با* $d \geq d_۰$*، وجود داشته باشد. آن‌گاه اگر* $y \in C_d$

$$\begin{aligned} h_Y(y\prime|y) &= \int_{X_۱} \pi_{Y|X}(y\prime|z) \pi_{X|Y}(z|y) \mu_۱(dz) \\ &\geq \int_{X_۱} \pi_{Y|X}(y\prime|z) \ell(z) \mu_۱(dz) = g(y\prime) \end{aligned}$$

اگرچه آن‌را به‌طور رسمی بیان نکردیم این از اثبات واضح است که اگر ما شرایط تحت این‌که P_X را به جای P_Y دوباره فرمول‌نویسی کنیم همان نتایج به‌دست می‌آید.

مثال ۶.۳.۳. مثال ۵.۸.۱ در نظر بگیرید. رابرتز و توییدی (۱۹۹۶) نشان دادند که متروپولیس قدم‌زدن تصادفی با توزیع هدف π نمی‌تواند ارگودیک هندسی باشد. اگر روش به‌روز رسانی مولفه به مولفه را در نظر بگیرید که توزیع هدف با توزیع شرطی مرتبط باشد، داریم $Y|X = x \sim N(۰, \frac{(۱+x^۲)^{-۱}}{۲})$ با محاسبه

$$\begin{aligned} \pi(x) &= \int e^{-x^۲} e^{-y^۲(۱+x^۲)} dy \\ &= e^{-x^۲} \int e^{-y^۲(۱+x^۲)} dy \\ &= e^{-x^۲} \int e^{-y^۲} \frac{۱}{۲\frac{(۱+x^۲)^{-۱}}{۲}} dy \\ &= e^{-x^۲} \sqrt{۲\pi \frac{(۱+x^۲)^{-۱}}{۲}} \left(\int \frac{۱}{\sqrt{۲\pi \frac{(۱+x^۲)^{-۱}}{۲}}} e^{-y^۲ \frac{۱}{۲\frac{(۱+x^۲)^{-۱}}{۲}}} dy \right) \\ &= e^{-x^۲} \sqrt{۲\pi \frac{(۱+x^۲)^{-۱}}{۲}} \end{aligned}$$

بنابراین

$$\begin{aligned}\pi(Y|X)=\frac{\pi(x,y)}{\pi(x)} &= \frac{e^{-x^{۲}}e^{-y^{۲}(۱+x^{۲})}}{e^{-x^{۲}}\sqrt{۲\pi\frac{(۱+x^{۲})^{-۱}}{۲}}}\\ &= \frac{۱}{\sqrt{۲\pi\frac{(۱+x^{۲})^{-۱}}{۲}}}e^{-y^{۲}(۱+x^{۲})}\\ &= \frac{۱}{\sqrt{۲\pi\frac{(۱+x^{۲})^{-۱}}{۲}}}e^{-y^{۲}\frac{۱}{۲\frac{(۱+x^{۲})^{-۱}}{۲}}}\end{aligned}$$

بدست می‌آید و به مراتب

$$X|Y=y\sim N(۰,\frac{(۱+y^{۲})^{-۱}}{۲}).$$

فورت و همکاران (۲۰۰۳)، ارگودیک هندسی و اسکن تصادفی یکنواخت متروپولیس قدم‌زدن تصادفی را ایجاد کردند، به این معنا که در هر مرحله یکی از مولفه‌ها با احتمال $\frac{۱}{۲}$ انتخاب می‌شود. در حالی‌که مرحله متروپولیس قدم‌زدن تصادفی برای دیگر مولفه ثابت باقی می‌ماند. ما می‌خواهیم نشان دهیم که نمونه‌گیر گیبز ارگودیک هندسی است. در این مثال اگر $W(y)=y^{۲}$ باشد آن‌گاه $P_YW(y)\leq\lambda W(y)+۰/۵$ برای هر $۰<\lambda<۱$.

$$\begin{aligned}h_Y(y\prime|y) &= \int\pi_{Y|X}(y\prime|x)\pi_{X|Y}(x|y)\pi(x)dx\\ &\geq \int\frac{۱}{\sqrt{\pi(x^{۲}+۱)^{-۱}}}e^{-y\prime^{۲}(x^{۲}+۱)}\frac{۱}{\sqrt{\pi}}e^{-x^{۲}(d+۱)}\sqrt{\pi(x^{۲}+۱)^{-۱}}e^{-x^{۲}}dx\\ &= \int\frac{۱}{\sqrt{\pi}}e^{-y\prime^{۲}}e^{-x^{۲}(y\prime^{۲}+۲+d)}dx\\ &= \frac{e^{-y\prime^{۲}}}{\sqrt{\pi}}\sqrt{\pi(y\prime^{۲}+۲+d)^{-۱}}\int\frac{۱}{\sqrt{\pi(y\prime^{۲}+۲+d)^{-۱}}}e^{-x^{۲}(y\prime^{۲}+۲+d)}dx\\ &= \frac{e^{-y\prime^{۲}}}{\sqrt{y\prime^{۲}+۲+d}}\\ &= g(y\prime)\end{aligned}$$

همچنین برای بررسی شرایط قضیه ۲.۳.۳ داریم که

$$\begin{aligned}
h_Y(y\prime|y) &= \int \pi_{Y|X}(y\prime|x)\pi_{X|Y}(x|y)\pi(x)dx \\
&= \int \frac{۱}{\sqrt{\pi(x^۲+۱)^{-۱}}}e^{-y\prime^۲(۱+x^۲)}\frac{۱}{\sqrt{\pi(y^۲+۱)^{-۱}}}e^{-x^۲(۱+y^۲)}\sqrt{\pi(x^۲+۱)^{-۱}}e^{-x^۲}dx \\
&= \frac{e^{-y\prime^۲}}{\sqrt{\pi(y^۲+۱)^{-۱}}}\int e^{-x^۲(y\prime^۲+۲+y^۲)}dx \\
&= \frac{\sqrt{\pi(y\prime^۲+y^۲+۲)^{-۱}}}{\sqrt{\pi(y^۲+۱)^{-۱}}}e^{-y\prime^۲} \\
&= \sqrt{\frac{y^۲+۱}{y\prime^۲+y^۲+۲}}e^{-y\prime^۲}
\end{aligned}$$

بنابراین

$$\begin{aligned}
P_Y W(y) &= \int W(y\prime)h_Y(y\prime|y)\pi(y\prime)dy\prime \\
&= \int y\prime^۲\sqrt{\frac{y^۲+۱}{y\prime^۲+y^۲+۲}}e^{-y\prime^۲}e^{-y\prime^۲}\sqrt{\pi(y\prime^۲+۱)^{-۱}}dy\prime \\
&= \sqrt{\pi(y^۲+۱)}\int\sqrt{\frac{y\prime^۴}{(y\prime^۲+۱)(y\prime^۲+y^۲+۱)}}e^{-۲y\prime^۲}dy\prime \\
&\leq \sqrt{\pi(y^۲+۱)}\int e^{-۲y\prime}dy\prime \\
&= \sqrt{\pi(y^۲+۱)}\sqrt{\frac{\pi}{۲}}\int\frac{۱}{\sqrt{\frac{\pi}{۲}}}e^{-۲y\prime^۲}dy\prime
\end{aligned}$$

علاوه بر این رابطه (۱۱.۳) اگر $y \in C_d$ باشد برای هر $d > ۰$ برقرار است. آن‌گاه

$$\begin{aligned}
\pi_{X|Y}(x|y) &= \frac{\sqrt{y^۲+۱}}{\sqrt{\pi}}e^{-x^۲(y^۲+۱)} \\
&\geq \frac{۱}{\sqrt{\pi}}e^{-x^۲(y^۲+۱)} \\
&\geq \pi^{-۰/۵}e^{-x^۲(d+۱)}
\end{aligned}$$

از این‌رو به‌وسیله قضیه ۲.۳.۳ ارگودیک هندسی بودن این مثال اثبات می‌شود.

اگرچه نمونه‌گیری‌های گیبز به‌خصوص برای مسایل آماری مناسب، ارگودیک هندسی بودنشان، اثبات شده است. تنها در داس و هابرت (۲۰۱۰) و هابرت و گیر (۱۹۹۸)، مثال‌هایی با بیش از دو مولفه ارایه شدند که، ارگودیک هندسی‌شان مورد بررسی قرار گرفته است. بنابراین قضیه ۲.۳.۳ می‌تواند، با نتایج موجود برای به‌دست آوردن ارگودیک هندسی دنباله تصادفی و نسخه‌های اسکن تصادفی بسیاری از نمونه‌گیری‌های گیبز، که ارگودیک هندسی بودنشان اثبات شده است، مرتبط باشد.

حالا فرض کنید قادر هستیم تا از $\pi_{X|Y}$ انتخاب نمونه کنیم اما به جای نمونه‌گیری از $\pi_{X|Y}$ یک مرحله متروپولیس–هستینگز، $g_۲$ با چگالی پیشنهادی، $P_۲$ جایگزین کنیم. این نتایج در یک نمونه‌گیری ترکیب، اغلب متروپولیس–هستینگز درون گیبز نامیده می‌شود، که دارای هسته مارکوف P_{HC} و Mtd

$$h_{HC}(x\prime, y\prime|x, y) = \pi_{X|Y}(x\prime|y)g_۲(y\prime|x\prime, y).$$

دنباله حاشیه‌ای Y با هسته P_Y مارکوفی است، و دارای Mtd

$$h_Y(y\prime|y) = \int \pi_{X|Y}(x|y)g_۲(y\prime|x, y)\mu_۱(dx) \qquad (۱۵.۳)$$

و چگالی مانای π_Y است، اما دنباله حاشیه‌ای X مارکوفی نیست. با این حال در قضیه ۴.۱ رابرت (۱۹۹۵) نشان داده شده است، که همگرایی دنباله X و Y همان نرخ، در نرم تغییرات کل است.

پس، فرض کنید $(X^{(۰)}, Y^{(۰)}) = (x, y)$ توزیع حاشیه‌ای $X^{(n)}$ با فضای اولیه $\tilde{P}^n_X((x, y), .)$ باشد بنابراین برای هر $n \geq ۱$

$$\|P^{n+۱}_Y(y, \cdot) - \varpi_Y(\cdot)\| \leq \|\tilde{P}^n_X((x, y), \cdot) - \varpi_X(\cdot)\| \leq \|P^n_Y(y, \cdot) - \varpi_Y(\cdot)\|$$

یک محاسبه ساده نشان می‌دهد که

$$\|P^{n+۱}_Y(y, \cdot) - \varpi_Y(\cdot)\| \leq \|P^n_{HC}((x, y), \cdot) - \varpi(\cdot)\|$$

این همچنین ساده است تا ببینید که دنباله Y برای P_{HC} بدون مقدار اولیه است، بنابراین P_Y ارگودیک هندسی است اگر و تنها اگر P_{HC} ارگودیک هندسی باشد (رابرتز و روزنتال (۲۰۰۱)). نتیجه بعدی ما به نرخ همگرایی P_Y و P_{HC} با نرخ همگرایی اسکن تصادفی زنجیر مرکب با هسته P_{RSH} و Mtd

$$h_{RSH}(x\prime, y\prime|x, y) = r\pi_{X|Y}(x\prime|y)\delta(y\prime - y) + (۱ - r)g_۲(y\prime|x, y)\delta(x\prime - x)$$

مربوط می‌شود. این ساده است که نشان دهیم P_{RSH} با توجه به ϖ برگشت‌پذیر است. توجه کنید که به‌وسیله (۲.۳) و (۱۵.۳) داریم که

$$P_Y\, W(Y) = \int_{X_۲} W(y\prime)h_Y(y\prime|y)\mu_۲(dy\prime).$$

قضیه ۷.۳.۳. *فرض کنید $W : X_۲ \longrightarrow \mathbb{R}^+$ و ثابت‌های $\lambda, b < \infty$ وجود داشته باشد آن‌گاه*

$$P_Y W(y) \leq \lambda W(y) + b \qquad (۱۶.۳)$$

فرض کنید $C_d = \{y : W(y) \leq d\}$، $g : X_۲ \longrightarrow \mathbb{R}^+$ و یک $d_۰ > ۰$، وجود داشته باشد به‌طوری‌که برای برخی $m \geq ۱$

$$h^m_Y(y\prime|y) \geq g(y\prime) \qquad \text{برای همه} \quad y \in C_d \text{ و } d \geq d_۰ \qquad (۱۷.۳)$$

آن‌گاه P_Y و P_{HC} ارگودیک هندسی هستند. علاوه بر این فرض کنید چگالی پیشنهادی $p_۲$ برای متروپولیس–هستینگز مرحله $g_۲$،

۱) *$p_۲(z|x, y) = p_۲(y|x, z)$ و $k < \infty$ وجود داشته باشد که $p_۲(z|x, y)/p_۲(z|x, u) \leq k$،*

۲) $p_۲(z|x, y) = p_۲(z|x)$

به‌دست آید، آن‌گاه P_{RSH} ارگودیک هندسی است.

با اثبات یک لم ابتدایی مربوط به $g_۲$ شروع می‌کنیم. Mtd به‌روز رسانی متروپولیس-هستینگز دارای چگالی هدف $\pi_{Y|X}(y|x)$، و چگالی پیشنهادی $p_۲(z|x,y)$ است. قرار دهید

$$\alpha(x,y,z) = ۱ \wedge \frac{\pi_{Y|X}(z|x)p_۲(y|x,z)}{\pi_{Y|X}(y|x)p_۲(z|x,y)}$$

و $\tilde{r}(x,y) = ۱ - \int p_۲(z|x,y)\alpha(x,y,z)\mu_۲(dz)$ همچنین داریم که

$$g_۲(z|x,y) = p_۲(z|x,y)\alpha(x,y,z) + \delta(z-y)\tilde{r}(x,y) \qquad (۱۸.۳)$$

لم ۸.۳.۳. *فرض کنید چگالی پیشنهادی $p_۲$ شرط ۱ یا ۲، از قضیه ۷.۳.۳ را دارا باشد. آن‌گاه*

$$p_۲(z|x,y)\alpha(x,y,z)p_۲(y|x,u)\alpha(x,u,y) \le kp_۲(z|x,u)p_۲(y|x,u)\alpha(x,,u,z) \qquad (۱۹.۳)$$

برهان. توجه کنید که

$$\begin{aligned}
p_۲(z|x,y)\alpha(x,y,z)p_۲(y|x,u)\alpha(x,u,y) &= p_۲(z|x,y)p_۲(y|x,u) \\
&\quad \times \left(۱ \wedge \frac{\pi_{Y|X}(z|x)p_۲(y|x,z)}{\pi_{Y|X}(y|x)p_۲(z|x,y)} \right) \\
&\quad \times \left(۱ \wedge \frac{\pi_{Y|X}(y|x)p_۲(u|x,y)}{\pi_{Y|X}(u|x)p_۲(z|x,y)} \right) \\
&\le p_۲(z|x,y)p_۲(y|x,u) \\
&\quad \times \left(۱ \wedge \frac{\pi_{Y|X}(z|x)p_۲(y|x,z)p_۲(u|x,y)}{\pi_{Y|X}(u|x)p_۲(z|x,y)p_۲(y|x,u)} \right)
\end{aligned}$$

اگر $p_۲(z|x,y) = p_۲(y|x,z)$ و $k < \infty$ وجود داشته باشد، آن‌گاه $p_۲(z|x,y)/p_۲(z|x,u) < k$ حاصل می‌شود. بنابراین

$$\begin{aligned}
&p_۲(z|x,y)p_۲(y|x,u)\left(۱ \wedge \frac{\pi_{Y|X}(z|x)p_۲(y|x,z)p_۲(u|x,y)}{\pi_{Y|X}(u|x)p_۲(z|x,y)p_۲(y|x,u)} \right) \\
&\quad = p_۲(z|x,y)p_۲(y|x,u)\left(۱ \wedge \frac{\pi_{Y|X}(z|x)}{\pi_{Y|X}(u|x)} \right) \\
&\quad = p_۲(z|x,y)p_۲(y|x,u)\alpha(x,u,z) \\
&\quad = p_۲(z|x,y)p_۲(y|x,u)\frac{p_۲(z|x,u)}{p_۲(z|x,u)}\alpha(x,u,z) \\
&\quad \le kp_۲(y|x,u)p_۲(z|x,u)\alpha(x,u,z).
\end{aligned}$$

اگر $p_۲(\cdot|x,y) = p_۲(\cdot|x)$، آن‌گاه

$$\begin{aligned}
&p_۲(z|x,y)p_۲(y|x,u)\left(۱ \wedge \frac{\pi_{Y|X}(z|x)p_۲(y|x,z)p_۲(u|x,y)}{\pi_{Y|X}(u|x)p_۲(z|x,y)p_۲(y|x,u)} \right) \\
&\quad = p_۲(z|x)p_۲(y|x)\left(۱ \wedge \frac{\pi_{Y|X}(z|x)}{\pi_{Y|X}(u|x)}\frac{p_۲(u|x)}{p_۲(z|x)} \right) \\
&\quad = p_۲(z|x,y)p_۲(y|x,u)\alpha(x,u,z) \\
&\quad = p_۲(z|x)p_۲(y|x)\alpha(x,u,z)
\end{aligned}$$

و از این‌رو رابطه (۱۹.۳) با $k = ۱$ برقرار است. □

P_{HC} و P_Y، m امین مرحله Mtd که با h^m_{HC} و h^m_Y نشان داده شده است را می‌پذیرند. لم زیر نشان می‌دهد که این دو Mtd با هم مرتبط هستند و از این روابط در اثبات قضیه استفاده خواهد شد.

لم ۹.۳.۳. *اگر $m \geq ۲$ باشد، آن‌گاه*

$$h^m_{HC}(x', y'|x, y) = \int_{x_۲} g_۲(y'|x', z)\pi_{X|Y}(x'|z)h^{m-۱}_Y(z|y)\mu_۲(dz)$$

برهان. فرض کنید $m = ۲$ باشد آن‌گاه

$$\begin{aligned} h^۲_{HC}(x', y'|x, y) &= \int_{x_۲}\int_{x_۱} \pi_{X|Y}(x'|\upsilon)g_۲(y'|x', \upsilon)\pi_{X|Y}(u|y)g_۲(\upsilon|u, y)\mu_۱(du)\mu_۲(d\upsilon) \\ &= \int_{x_۲} \pi_{X|Y}(x'|\upsilon)g_۲(y'|x', \upsilon)\int_{x_۱} \pi_{X|Y}(u|y)g_۲(\upsilon|u, y)\mu_۱(du)\mu_۲(d\upsilon) \\ &= \int_{x_۲} \pi_{X|Y}(x'|\upsilon)g_۲(y'|x', \upsilon)h_Y(\upsilon|y)\mu_۲(d\upsilon) \end{aligned}$$

با یک محاسبه مشابه نتیجه برقرار است و اثبات کامل می‌شود:

$$\begin{aligned} h^{m+۱}_{HC}(x', y'|x, y) &= \int_{x_۲}\int_{x_۱} h(x', y'|u, \upsilon)h^m(u, \upsilon|x, y)\mu_۱(du)\mu_۲(d\upsilon) \\ &= \int_{x_۲}\int_{x_۱}\int_{x_۲} \pi_{X|Y}(x'|\upsilon)g_۲(\upsilon|u, z)g_۲(\upsilon|u, z)\pi_{X|Y}(u|z) \\ &\qquad h^{m-۱}_Y(z|y)\mu_۲(dz)\mu_۱(du)\mu_۲(d\upsilon) \\ &= \int_{x_۲}\int_{x_۲} \pi_{X|Y}(x'|\upsilon)g_۲(y'|x', \upsilon)h_y(\upsilon|z)h^{m-۱}_Y(z|y)\mu_۲(dz)\mu_۲(d\upsilon) \\ &= \int_{x_۲} \pi_{X|Y}((x'|\upsilon)g_۲(y'|x', \upsilon)h^m_Y(\upsilon|y)\mu_۲(dz)\mu_۲(d\upsilon). \end{aligned}$$

□

برهان. وجه تشابه با ابتدای اثبات قضیه ۲.۳.۳ این است که، می‌خواهیم نشان دهیم که P_Y و P_{HC} ارگویک هندسی هستند. فرض (۱۶.۳) و تعریف $G(x, y) = \int_{X_۲} W(z)g_۲(z|x, y)\mu_۲(dz)$ را به‌خاطر بیاورید. فرض کنید احتمال انتخاب $r > (k + ۱)/(k + ۲ - \beta)$،

$$\frac{۱ - r}{۱ - (۱ - r)(k + ۲)} < \upsilon < \frac{۱ - r}{r\lambda}$$

و $V(x, y) = \upsilon G(x, y) + W(y)$. آن‌گاه

$$\begin{aligned} P_{RSH}V(x, y) &= r\int V(x', y')\pi_{X|Y}(x'|y)\delta(y' - y)\mu_۱(dx')\mu_۲(dy') \\ &\quad +(۱ - r)\int V(x', y')g_۲(y'|x, y)\delta(x' - x)\mu_۱(dx')\mu_۲(dy') \\ &= r\int V(x', y)\pi_{X|Y}(x'|y)\mu_۱(dx') \\ &\quad +(۱ - r)\int V(x, y')g_۲(y'|x, y)\mu_۲(dy') \end{aligned}$$

$$\begin{aligned}
&= r\int(\upsilon G(x',y)+W(y))\pi_{X|Y}(x'|y)\mu_۱(dx')\\
&\quad +(۱-r)\int(\upsilon G(x,y')+W(y'))g_۲(y'|x,y)\mu_۲(dy')\\
&= r\upsilon\int G(x',y)\pi_{X|Y}(x'|y)\mu_۱(dx')+rW(y)\\
&\quad +(۱-r)\upsilon\int G(x,y')g_۲(y'|x,y)\mu_۲(dy')\\
&\quad +(۱-r)\int W(y')g_۲(y'|x,y)\mu_۲(dy')\\
&= r\upsilon\int G(x',y)\pi_{X|Y}((x'|y)\mu_۱(dx')\\
&\quad +(۱-r)\upsilon\int G(x,y')g_۲(y'|x,y)\mu_۲(dy')\\
&\quad +rW(y)+(۱-r)G(x,y)
\end{aligned}$$

به خاطر بیاورید که $h_Y(\cdot|y)$، Mtd برای دنباله Y است و توجه کنید که به‌وسیله رابطه (۱۶.۳) داریم

$$\begin{aligned}
\int G(x',y)\pi_{X|Y}(x'|y)\mu_۱(dx') &= \int\int W(z)g_۲(z|x',y)\pi_{X|Y}(x'|y)\mu_۱(dx')\mu_۲(dz)\\
&= \int W(z)h_Y(z|y)\mu_۲(dz)\\
&\le \lambda W(y)+b
\end{aligned}$$

بنابراین

$$P_{RSH}V(x,y)\le(۱-r)\upsilon\int G(x,y')g_۲(y'|x,y)\mu_۲(dy')+r(۱+\upsilon\lambda)W(y)+(۱-r)G(x,y)+r\upsilon b.$$

حال به‌وسیله (۱۸.۳) می‌توان نوشت:

$$\begin{aligned}
\int G(x,y')g_۲(y'|x,y)\mu_۲(dy') &= \int G(x,y')[p_۲(y'|x,y)\alpha(x,y,y')\\
&\quad +\delta(y'-y)\tilde{r}(x,y)]\mu_۲(dy')\\
&= \int G(x,y')p_۲(y'|x,y)\alpha(x,y,y')\mu_۲(dy')+G(x,y)\tilde{r}(x,y)\\
&\le \int G(x,y')p_۲(y'|x,y)\alpha(x,y,y')\mu_۲(dy')+G(x,y).
\end{aligned}$$

اولین عبارت را $A(x,y)$ تعریف کنید و همچنین با رابطه (۱۹.۳) داریم

$$\begin{aligned}
A(x,y) &= \int G(x,y')p_۲(y'|x,y)\alpha(x,y,y')\mu_۲(dy')\\
&= \int\int W(z)g_۲(z|x,y')p_۲(y'|x,y)\alpha(x,y,y')\mu_۲(dz)\mu_۲(dy')\\
&= \int\int W(z)p_۲(z|x,y')\alpha(x,y',z)p_۲(y'|x,y)\\
&\quad \alpha(x,y',z)p_۲(y'|x,y)\alpha(x,y,y')\mu_۲(dz)\mu_۲(dy')\\
&\quad +\int W(y')\tilde{r}(x,y')p_۲(y'|x,y)\alpha(x,y,y')\mu_۲(dy')
\end{aligned}$$

$$\begin{aligned} &\leq K\int\int W(z)p_2(z|x,y)p_2(y'|x,y)\alpha(x,y,z)\mu_2(dz)\mu_2(dy') \\ &\quad + \int W(y')g_2(y'|x,y)\mu_2(dy') \\ &= K\int W(z)p_2(z|x,y)\alpha(x,y,z)\mu_2(dz) + G(x,y) \\ &\leq (K+1)G(x,y). \end{aligned}$$

با قرار دادن نتایج بالا کنار یکدیگر داریم

$$P_{RSH}V(x,y) \leq \frac{(1-r)[1+\nu(K+2)]}{\nu}\nu G(x,y) + r(1+\nu\lambda)W(y) + r\nu b$$

فرضیات r و ν وجود یک ثابت β را تضمین می‌کند به‌طوری که

$$\max\left\{\frac{(1-r)[1+\upsilon(K+2)]}{\upsilon}, r(1+\upsilon\lambda)\right\} \leq \beta < 1$$

از این‌رو

$$P_{RSH}V(x,y) \leq \beta V(x,y) + r\upsilon b.$$

مجموعه $U = 1+V$، $\gamma = (\beta+1)/2$ و $C = \{(x,y) : U(x,y) \leq r\nu b/(1-\gamma)\}$. آن‌گاه به‌وسیله لم ۳.۳.۳ داریم

$$P_{RSH}U(x,y) \leq \gamma U(x,y) + r\upsilon b I_C(x,y).$$

حالا ما باید بیان کنیم که C کوچک است. توجه کنید که $P^{2m}_{RSH}((x,y),\cdot) \geq [r(1-r)]^m P^m_{HC}((x,y),\cdot)$ است، از این‌رو کافیست تا یک شرط کوچک‌سازی برای P^m_{HC} ایجاد کنیم.

فرض کنید $C_d = \{y : W(y) \leq d\}$. کافیست که اگر $d \geq r\nu b/(1-\gamma)$، یک $X_1 \times C_d$ کوچک، زمانی‌که $C \subseteq X_1 \times C_d$، ایجاد کنیم. توجه کنید که ما مجبوریم که تنها، حالتی‌که $m \geq 2$ است را در نظر بگیریم. چون‌که اگر رابطه (۱۷.۳) با $m = 1$ برقرار باشد، آن‌گاه

$$h^2(y'|y) = \int_{x_2} h_Y(y'|z)\mu_2(dz) \geq \int_{x_2} h_Y(y'|z)g(z)\mu_2(dz) := g*(y')$$

از این‌رو رابطه (۱۷.۳) با $m = 2$ برقرار است. با لم ۸.۳.۳ و رابطه (۱۷.۳) اگر $y \in C_d$ ، آن‌گاه برای برخی $m \geq 2$

$$\begin{aligned} h^{m+1}_{HC}(x',y'|x,y) &= \int_{x_2} g_2(y'|x',z)\pi_{X|Y}(x'|z)h^m_Y(z|y)\mu_2(dz) \\ &\geq \int_{x_2} g_2(y'|x',z)\pi_{X|Y}(x'|z)g(z)\mu_2(dz) \\ &= \epsilon q(x',y') \end{aligned}$$

با

$$\epsilon = \int_{x_1}\int_{x_2} \pi_{X|Y}(x'|z)g(z)\mu_2(dz)\mu_1(dx')$$

$$q(x',y') = \epsilon^{-1}\int_{x_2} g_2(y'|x',z)\pi_{X|Y}(x'|z)g(z)\mu_2(dz)$$

بیان می‌شود که $X_1 \times C_d$ برای P_{HC}^{m+1}، و همچنین $P_{RSHC}^{2(m+1)}$ کوچک است. همچنین نشان دادیم که $P_{HC}^{2(m+1)}$، برای $r > (K+1)/(K+2-\lambda)$، ارگودیک هندسی است. با رجوع به قضیه ۱.۳.۳ اثبات کامل می‌شود. □

نتیجه ۱۰.۳.۳. *با تنظیمات نمونه‌گیر گیبز یک شرط کافی و ساده برای رابطه (۱۷.۳) وجود دارد؛ فرض کنید یک تابع نامنفی ℓ، وجود داشته باشد به‌طوری‌که برای $y \in C_d$*

$$\pi_{X|Y}(x|y)g_2(z|x,y) \geq \ell(x,z)$$

در این حالت اگر $y \in C_d$ باشد آن‌گاه

$$h_Y(y\prime|y) = \int_{X_1} \pi_{Y|X}(y\prime|x)g_2(y\prime|x,y)\mu_1(dx) \geq \int_{X_1} \ell(x,y\prime)\mu_1(dx) = g(y\prime).$$

۲.۳.۳ نمونه‌گیری‌های گیبز برای مدل آمیخته خطی بیزی

فرض کنید Y نشان‌دهنده یک بردار پاسخ $N \times 1$، β یک بردار $p \times 1$ از ضرایب رگرسیونی، u یک بردار $K \times 1$ از اثرات تصادفی، X یک ماتریس طرح $N \times P$ معلوم، و Z یک ماتریس $N \times K$ معلوم، باشد. همچنین فرض کنید به ازای $r, s, t \in \{1, 2, \dots\}$

$$Y|\beta, u, \lambda_R, \lambda_D \sim N_N(X\beta + Zu, \lambda_R^{-1} I_N),$$

$$\beta|u, \lambda_R, \lambda_D \sim \sum_{i=1}^{r} \eta_i N_p(b_i, B^{-1}),$$

$$u|\lambda_R, \lambda_D \sim N_k(0, \lambda_D^{-1} I_k),$$

$$\lambda_R \sim \sum_{j=1}^{s} \phi_j Gamma(r_{j1}, r_{j2}),$$

$$\lambda_D \sim \sum_{\ell=1}^{t} \psi_\ell Gamma(d_{\ell 1}, d_{\ell 2}),$$

که پارامترهای آمیخته η_i، ϕ_j و ψ_ℓ ثابت‌های نامنفی معلوم هستند که

$$\sum_{i=1}^{r} \eta_i = \sum_{j=1}^{s} \phi_j = \sum_{\ell=1}^{t} \psi_\ell = 1.$$

که در آن $W \sim Gamma(a,b)$، دارای چگالی پیشنهادی به‌صورت $\omega^{a-1} e^{-b\omega} I(\omega > 0)$. در نهایت همچنین فرض می‌کنیم $X^T Z = 0, b_i \in \mathbb{R}$ و ماتریس معین مثبت B معلوم است و ابرپارامترهای r_{j1}، r_{j2}، $d_{\ell 1}$ و $d_{\ell 2}$ مثبت هستند.

فرض کنید $\xi = (u^T, \beta^T)^T$ و $\lambda = (\lambda_R, \lambda_D)^T$. آن‌گاه چگالی پسین به‌وسیله

$$\pi(\xi, \lambda|y) \propto f(y|\xi, \lambda) f(\xi|\lambda) f(\lambda),$$

مشخص شده است که y داده مشاهده شده و f نشان‌دهنده یک چگالی کلی است. ساده است تا توزیع‌های شرطی $\lambda|\xi,y$ و $\xi|\lambda,y$ را به‌دست آوریم که در این‌جا توصیف شده است. فرض کنید

$$\nu_۱(\xi)=(y-X\beta-Zu)^T(y-X\beta-Zu)$$

و

$$\nu_۲(\xi)=u^Tu.$$

آن‌گاه توزیع $\lambda|\xi,y$ دارای چگالی

$$f(\lambda|\xi,y)=\sum_{j=۱}^{s}\sum_{\ell=۱}^{t}\phi_j\psi_\ell f_{۱j}(\lambda_R|\xi,y)f_{۲\ell}(\lambda_D|\xi,y)$$

است. برای سادگی، اثبات را زمانی‌که $r=s=۱$، در نظر می‌گیریم. ابتدا برای پارامترهای دقیق با ابرپارامترهای $r_۱$ و $r_۲$ و $d_۱$ و $d_۲$ داریم

$$\begin{aligned}\pi(\lambda_R|\xi,y) &\propto \pi(y|\lambda_R,\xi)f(\lambda_R)\\ &\propto \lambda_R^{N/۲}.exp\{-۰/۵\lambda_R\nu_۱(\xi)\}.\lambda_R^{r_۱-۱}exp\{-r_۲\lambda_R\}\\ &\propto \lambda_R^{(r_۱+N/۲)-۱}.exp\{-\lambda_R(r_۲+۰/۵\nu_۱(\xi))\}\end{aligned}$$

و

$$\begin{aligned}\pi(\lambda_D|\xi,y) &\propto \pi(u|\lambda_D,\xi)f(\lambda_D)\\ &\propto \lambda_D^{k/۲}.exp\{-۰/۵\lambda_D\nu_۲(\xi)\}.\lambda_D^{d_۱-۱}exp\{-d_۲\lambda_D\}\\ &\propto \lambda_D^{(d_۱+k/۲)-۱}.exp\{-\lambda_D(d_۲+۰/۵\nu_۲(\xi))\}\end{aligned}$$

که $f_{۱j}(.|\xi,y)$ یک چگالی $Gamma(r_{j۱}+N/۲,r_{j۲}+\nu_۱(\xi)/۲)$ و $f_{۲\ell}(.|\xi,y)$ یک چگالی $Gamma(d_{\ell۱}+k/۲,d_{\ell۲}+\nu_۲(\xi)/۲)$ را نشان می‌دهد. از طرفی توجه کنید که برای هر $W\sim N_m(\mu_w,\sum_w^{-۱})$ چگالی به‌صورت

$$\pi(w)\propto exp\{-۰/۵(w-\mu_W)^T\sum_W(w-\mu_W)\}\propto exp\{-۰/۵w^T\sum_W w+w^T\sum_W\mu_W\}$$

آن‌گاه برای شرطی کامل ξ

$$\begin{aligned}\pi(\xi|\lambda,y) &\propto \pi(y|\beta,u,\lambda).\pi(\beta|u,\lambda).\pi(u|\lambda)\\ &\propto exp\{-۰/۵\lambda_R(y-X\beta-Zu)^T(y-X\beta-Zu)\}\\ &\quad .exp\{-۰/۵[(\beta-b)^TB(\beta-b)+\lambda_Du^Tu]\}\\ &\propto exp\{-۰/۵\lambda_R[y^Ty-y^TX\beta-y^TZu-\beta^TX^Ty\\ &\quad +\beta^TX^TX\beta+\beta^TX^TZu-u^TZ^Ty+u^TZ^TX\beta+u^TZ^TZu]\}\\ &\quad .exp\{-۰/۵[\beta^TB\beta-\beta^TBb-b^TB\beta+b^TBb+\lambda_Du^Tu]\}\\ &\propto exp\{-۰/۵[\beta^T(\lambda_RX^TX+B)\beta+u^T(Z^TZ+\lambda_D)u]\}\\ &\quad exp\{-۰/۵[-۲\beta^T(\lambda_RX^Ty+Bb)-۲\lambda_Ru^TZ^Ty+۲\lambda_R\beta^TX^TZu]\}\end{aligned}$$

آن‌گاه توزیع $\xi|\lambda,y\sim\sum_{i=1}^{r}\eta_i N(m_0,\sum^{-1})$ با $X^TZ=0$، حاصل می‌شود که

$$\sum^{-1}=\begin{pmatrix}(\lambda_R Z^TZ+\lambda_D I_k)^{-1} & 0\\ 0 & (\lambda_R X^TX+B)^{-1}\end{pmatrix}$$

و

$$m_0=\begin{pmatrix}\lambda_R(\lambda_R Z^TZ+\lambda_D I_k)^{-1}Z^Ty\\ (\lambda_R X^TX+B)^{-1}(\lambda_R X^Ty+Bb)\end{pmatrix}.$$

اجرا کردن استراتژی‌های نمونه‌گیری گیبز ساده است. برای مثال نمونه‌گیری گیبز را در نظر بگیرید که، به‌روز رسانی ξ به‌وسیله λ با داشتن Mtd به‌صورت [به‌خاطر بیاورید رابطه (۸.۳)]

$$h_{GS}(\xi',\lambda'|\xi,\lambda)=f(\xi'|\lambda,y)f(\lambda'|\xi',y).$$

تعریف می‌شود. به‌طور مشابه می‌توانیم شرطی‌های کامل و دستورالعمل‌های سریع‌تر برای ساخت Mtd زنجیرهای مارکوف مربوطه که توصیف شده است، به‌کار ببریم و نتایج h_ξ، h_λ، $\tilde{h}_{GS}$، h_{RSGS}، h_{RQGS} را مشاهده کنیم. جانسون و جونز (۲۰۱۰) رابطه (۱۰.۳) و (۱۱.۳) را برای دنباله حاشیه‌ایی ξ با داشتن Mtd ، h_ξ بیان کردند، و از این‌رو ما می‌توانیم به قضیه ۲.۳.۳ برای ایجاد ارگودیک هندسی نمونه‌گیری‌های گیبز متوسل شویم.

فرض کنید x_i و z_i به ترتیب i امین سطر از X و Z باشند و تعریف کنید

$$G_i(\lambda)=\sum_{m=1}^{N}[E_i(y_m-x_m\beta-z_mu|\lambda,y)]^2+\sum_{m=1}^{k}[E_i(u_m|\lambda,y)]^2$$

که E_i نشان‌دهنده امید ریاضی با توجه به توزیع $N_{k+p}(m_i,\sum^{-1})$ است.

قضیه ۱۱.۳.۳. *فرض کنید مقدار $K<\infty$ وجود داشته باشد به‌طوری‌که $G_i(\lambda)\le K$. اگر برای همه $j\in\{1,\dots,s\}$ و $\ell\in\{1,\dots,t\}$*

$$r_{j1}>0\vee\frac{1}{2}[\sum_{i=1}^{N}z_i(Z^TZ)^{-1}z_i^T-N+2]$$

و $d_{\ell 1}>1$ باشد، آن‌گاه زنجیرهای حاشیه‌ایی λ و ξ و GS ارگودیک هندسی هستند از این‌رو $RSGS$ و $RQGS$ نیز ارگودیک هندسی هستند.

برهان. برای اثبات این قضیه می‌توانید به جانسون (۲۰۰۹) مراجعه کنید. □

جانسون و جونز (۲۰۱۰)، شرایطی دیگر تحت این‌که GS ارگودیک هندسی باشد، را فراهم کردند و درباره‌ی شرایط ارگودیک هندسی بودن RSGS و RQGS بحثی نمی‌کنند. همچنین می‌توانید به رومن (۲۰۱۲) برای ایجاد برخی تغییرات ایجاد شده و بهتر شدن نتایج، جانسون و جونز (۲۰۱۰)، مراجعه کنید. در فصل ۴، حالت خاصی از مدل‌مان را در نظر می‌گیریم و GS و RSGS و RQGS و یک نمونه متروپولیس شرطی کامل را با یکدیگر، مورد مقایسه تجربی قرار می‌دهیم.

فصل ۴

مطالعه شبیه‌سازی

۱.۴ مقدمه

در این فصل، نتایج حاصل از فصل‌های گذشته را در قالب یک مطالعه شبیه‌سازی مورد بررسی قرار می‌دهیم. ما نتایج را، برای دو مثال واقعی که شامل، یک مدل سلسله مراتبی آمیخته خطی و برآورد ماکسیمم درستنمایی برای مدل آمیخته، نشان می‌دهیم که، در این دو مثال، مقایسه تجربی بین نمونه‌ها با به‌کارگیری به‌روز رسانی مولفه به مولفه در مقابل به‌روز رسانی هم‌زمان همه متغیرها صورت گرفته است. نتایج، حاکی از کارایی مطلوب‌تر روش به‌روز رسانی مولفه به مولفه، نسبت به روش به‌روز رسانی هم‌زمان همه متغیرها می‌باشد.

با توجه به تحقیقاتی که توسط جانسون و جونز (۲۰۱۰)، کول و همکاران (۲۰۰۱)، کافو، جیک و جونز (۲۰۰۵)، مک‌کالاک (۱۹۹۷)، لی و همکاران (۲۰۱۳)، جونز و هوبرت (۲۰۰۱)، جونز و همکاران (۲۰۰۶)، نیس (۲۰۱۳) صورت گرفته، عملکرد تجربی نمونه‌های مولفه به مولفه از ضمانت بیش‌تری برخوردار هستند. این مثال‌ها بر اساس تنظیمات معرفی شده در بخش‌های ۲.۲.۳ و ۲.۳.۳ بیان می‌شوند. در این دو مثال، علاوه بر خلاصه عددی بر اساس معیارهای تعریف شده در فصل اول، که این معیارها، یک تصویر معقولی از عملکرد تجربی الگوریتم‌های مختلف را بیان می‌کنند، از نظر خلاصه گرافیکی مانند نمودار اثر نیز مورد مطالعه قرار می‌گیرند.

۲.۴ مدل آمیخته خطی بیزی

یک نسخه بیزی، مدلی با عرض از مبدا تصادفی متعادل شده، برای K مورد، و $m \geq ۲$ مشاهدات برای هر موضوعی در نظر بگیرید. فرض کنید $y_i = (y_{i۱}, \ldots, y_{im})^T$ داده‌ای برای i امین موضوع باشد و $Y = (y_۱^T, \ldots, y_k^T)^T$ به‌طور کلی نشان‌دهنده بردار پاسخ $N \times ۱$، باشد که $N = km$. علاوه بر این، فرض کنید $u = (u_۱, \ldots, u_k)^T$ یک برداری از اثرات موضوع و X یک ماتریس طرح $N \times P$ پر رتبه ستونی مرتبط با β باشد که β برداری $P \times ۱$ از ضرایب رگرسیونی است. آن‌گاه اولین سطح از سلسله

مراتب

$$Y|\beta, u, \lambda_R, \lambda_D \sim N_N(X\beta + Zu, \lambda_R^{-1} I_N)$$

برای $Z = I_k \otimes 1_m$ که $\otimes$ نشان‌دهنده عملگر کرونکر و 1_m یک برداری $m \times 1$ از یک‌هاست. مرحله بعدی مدل سلسله مراتبی

$$\beta|\lambda_R, \lambda_D \sim N_p(b, B^{-1})$$

و

$$u|\lambda_R, \lambda_D \sim N_k(0, \lambda_D^{-1} I_k)$$

برای $b \in \mathbb{R}^b$ معلوم، و ماتریس B معین مثبت است. در نهایت

$$\lambda_R \sim Gamma(r_1, r_2) \quad و \quad \lambda_D \sim Gamma(d_1, d_2)$$

که r_1 و r_2 و d_1 و d_2 مثبت هستند.

سلسله مراتبی حالت خاصی از مدل خطی بیزی در حالت کلی، که در بخش ۲.۳.۳ تعریف شده است که اگر $d_1 > 1$ باشد، از قضیه ۱۱.۳.۳ پیروی می‌کند و آنگاه الگوریتم‌های GS و RQGS و RSGS ارگودیک هندسی هستند. ما در حال حاضر یک مقایسه تجربی از الگوریتم‌های GS و RQGS یکنواخت و RSGS یکنواخت ارایه می‌دهیم. همچنین سه نمونه‌گیر گیبز را با متروپلیس قدم‌زدن تصادفی شرطی کامل مقایسه می‌کنیم. در مقایسه‌مان روی برآورد امید ریاضی پسین β که $E(\beta|y)$، است تمرکز می‌کنیم.

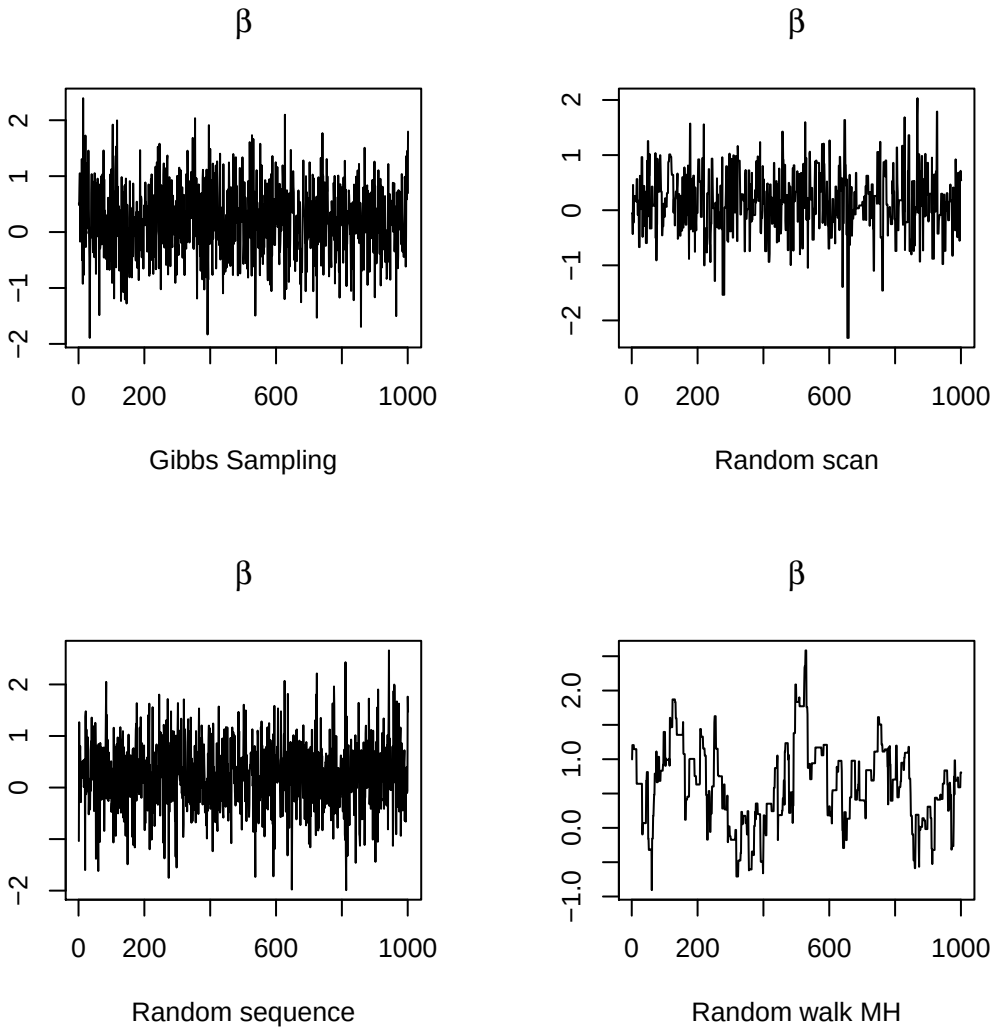
متروپلیس قدم‌زدن تصادفی RW، از یک توزیع پیشنهادی نرمال چند متغیره با یک ماتریس کواریانس قطری، که روی مقدار جاری زنجیر متمرکز شده است، استفاده می‌کند. عناصر قطری معادل را با $\hat{\sum}^2$ برآورد می‌کنیم که $\hat{\sum}$ یک برآوردی از ماتریس کواریانس پسین به‌دست آمده، از یک اجرای مستقل GS با 10^5 تکرار است. همچنین برای تنظیمات توصیف شده در زیر، اطلاع از ارگودیک هندسی بودن یا نبودن زنجیر مارکوف RW نداریم.

داده (مقادیر y) را تحت تنظیمات زیر شبیه سازی کردیم: $k = 10$ و $m = 5$ و $p = 1$ و $X = (x_1^T, \ldots, x_{10}^T)^T$، از طرفی برای همه iها، $x_i^T = (-0/5, -0/25, 0, 0/25, 0/5)$ با $b = 0$ و $B^{-1} = 0/1$ و $r_1 = r_2 = d_1 = d_2 = 2$ در نظر گرفته شده است.

فرض کنید ماهیت واقعی این داده نامعلوم است. چهار زنجیر مارکوف تحت ابرپارامتر تنظیم شده با $b = 0$ و $B^{-1} = 0/1$ و $r_1 = r_2 = d_1 = d_2 = 3$ شبیه‌سازی می‌کنیم. در نهایت تمام زنجیرها با پیشین $(\beta^{(0)}, u^{(0)}, \lambda_R^{(0)}, \lambda_D^{(0)}) = (0, 0_k, 1, 1)$، که 0_k برداری $k \times 1$ از صفرها هستند، شروع می‌شوند.

از آن‌جایی که $E[\beta^4|y] < \infty$، ارگودیک هندسی بودن GS ، RQGS و RSGS، تضمین کننده قضیه حد مرکزی برای خطای مونت کارلو $\bar{\beta}_n - E(\beta|y)$، با واریانس توزیع مجانبی، σ_β^2 می‌باشد. هر یک از چهار الگوریتم GS، RQGS ، RSGS و RW برای 10^5 تکرار، به‌طور مستقل اجرا می‌شوند. نمودار اثر β ، در شکل ۱.۴ نشان می‌دهد که آمیختگی نمونه‌گیر گیبز از RW به‌طور قابل ملاحظه‌ایی سریع‌تر است در حالی‌که RQGS و GS به نظر می‌رسد از RSGS کارآمدتر باشند.

تفاوت در نمودارهای اثر بین چهار زنجیر مارکوف شبیه‌سازی، در نصف فاصله عرضی منعکس شده است. برای نمونه با اندازه‌های یکسان، فاصله اطمینان به‌دست آمده از الگوریتم RW با فاصله اطمینان



شکل ۱.۴: نمودار اثر β در مدل آمیخته خطی بیزی به عنوان مثالی از بخش ۲.۴

الگوریتم RQGS یکسان است، اما نسبت به GS و RSGS طول فاصله اطمینان حدود شش برابر بزرگ‌تر است.

بزرگ بودن فاصله اطمینان به این معناست که، به تعداد تکرارهای بیش‌تری برای همگرایی نیاز داریم که هم زمان‌بر است و هم میزان دقت کاهش پیدا می‌کند. بنابراین با این خروجی به این نتیجه می‌رسیم که آمیختگی نمونه‌گیری گیبز از RSGS سریع‌تر است و همچنین دارای پهنای فاصله اطمینان کم‌تری نسبت به سایر الگوریتم‌هاست.

علاوه بر این، معیار ACT نشان‌دهنده کارایی زنجیر مارکوف است، هر قدر میزان این معیار کم‌تر باشد به معنای کارایی به‌تر الگوریتم است. با توجه به بالاتر بودن معیار ACT برای الگوریتم RW، کارایی نسبی آن نسبت به بقیه الگوریتم‌ها کم‌تر است و همچنین بر اساس این معیار الگوریتم‌های RQGS و RSGS کارایی یکسانی دارند، بنابراین الگوریتم GS با کم‌ترین ACT دارای به‌ترین کارایی است. البته

قابل ذکر است که تفاوت زیادی بین معیار ACT الگوریتم‌های RSGS و RQGS و GS وجود ندارد.

معیار بعدی که مورد ارزیابی قرار می‌گیرد، MSE است، این معیار برای سنجش خطای برآورد به‌کار می‌رود. هرقدر میزان این معیار کم‌تر باشد دقت الگوریتم بیش‌تر و دارای خطای کم‌تری است. به منظور برآورد MSE برآوردهای مونت کارلو، $m = ۵۰۰$ تکرار مستقل از RW، RSGS و GS برای هرکدام با $n = ۱۰^۵$ تکرار شبیه‌سازی کرده‌ایم و $\bar{\beta}^*$ یک برآورد از $E(\beta|y)$ تفسیر شده است، که از $۱۰^۵$ تکرار زنجیر RW به‌دست می‌آید.

چون علاقه‌مند به برآورد میانگین پسین $E(\beta|y)$ هستیم، برای مقایسه کیفیت برآوردهای GS و RQGS و RSGS و RW تجزیه و تحلیل مربوط به MSE را انجام می‌دهیم. برآورد نسبت MSE با GS مرتبط است

$$\frac{\widehat{MSE}(\bar{\beta}_{n,*})}{\widehat{MSE}(\bar{\beta}_{n,GS})}$$

که در جدول ۱.۴ همراه با خطای استاندارد نشان داده شده است. توجه داشته باشید که هر قدر میزان این نسبت به یک نزدیک باشد به معنای کارایی بالاتر GS نسبت به برآوردگرهای دیگر است. بر این اساس کارایی RW حدود دو برابر کم‌تر از GS است.

ESEJD فاصله اقلیدسی بین پرش‌ها را محاسبه می‌کند و هر قدر این معیار کم‌تر باشد پرش‌ها به هم نزدیک‌تر هستند و بنابراین همگرایی دیرتر رخ می‌دهد. ESEJD، بر اساس $m = ۵۰۰$ تکرار مستقل از الگوریتم‌های RW، RSGS، RQGS و GS برای $n = ۱۰^۵$ در هر تکرار برآورد می‌شود. برآورد همراه با خطای استاندارد در جدول ۱.۴ گزارش شده است، که با توجه به خروجی داده‌ها RW دارای ESEJD کم‌تری است بنابراین همگرایی در آن دیرتر اتفاق می افتد.

جدول ۱.۴: نتایج حاصل از مثال بخش ۲.۴، مدل آمیخته خطی بیزی.

Algorithm	$\bar{\beta}_n$	$\hat{\sigma}^۲_\beta$	$t_*\hat{\sigma}_\beta/\sqrt{n}$	ACT	$MSE\,ratio$	$\widehat{ESEJD}$
RW	۰٫۴۹	$۱٫۰۸ \times ۱۰^{-۶}$	$۶٫۴۴۲ \times ۱۰^{-۵}$	۵٫۴۸	۲٫۰۰۲ (۰٫۰۷۲۷)	۰٫۰۰۱۹ ($۱٫۰۰۲۹ \times ۱۰^{-۵}$)
RSGS	۱٫۲۸	$۳٫۸۷ \times ۱۰^{-۶}$	$۱٫۲۱۹ \times ۱۰^{-۵}$	۱٫۷۶	۱٫۰۰۶ (۰٫۰۷۸۴)	۰٫۳۷۷ (۰٫۰۰۰۴۸)
RQGS	۱٫۳۹	$۹٫۸۴ \times ۱۰^{-۶}$	$۶٫۱۴۹ \times ۱۰^{-۵}$	۱٫۱۳	۰٫۸۹۸ (۰٫۰۵۲)	۰٫۹۳۳ (۰٫۰۰۰۸۲)
GS	۱٫۲۷	$۳٫۸۴ \times ۱۰^{-۶}$	$۱٫۲۱۴ \times ۱۰^{-۵}$	۰٫۹۷	۱ (۰)	۰٫۷۵۵ (۰٫۰۰۰۶۲)

در جدول فوق برآورد $E(\beta|y)$، $\sigma^۲_\beta$، $Var(\beta|y)$، براساس $n = ۱۰^۵$، است. ستون میانی جدول، نشان‌دهنده نصف پهنای فاصله اطمینان ۹۵ درصدی با $t_* = ۱٫۹۶۰$ می‌باشد. همچنین معیارهای ACT، MSE ratio و برآورد ESEJD به همراه خطای استاندارد در داخل پرانتز نشان داده شده است.

۳.۴ مدل آمیخته لجستیک نرمال

حالت خاصی از مدل آمیخته تعریف شده در بخش ۲.۲.۳ را در زیر در نظر بگیرید. فرض کنید شرط روی $U = u$ باشد و مشاهدات Y_{ij} به‌طور مستقل از توزیع برنولی (p_{ij}) پیروی می‌کنند، که برای $j = ۱, \ldots, m_i$ و $i = ۱, \ldots, k$، $Logit(p_{ij}) = \beta x_{ij} + u_i$، که x_{ij} متغیرهای کمکی هستند. فرض

کنید اثرات تصادفی $U_1, \ldots, U_k$ یکسان و مستقل و هم‌توزیع از $N(0, \sigma^2)$، و همچنین پارامترهای $\theta = (\beta, \sigma^2)$ ثابت باشند. چگالی هدف

$$h(u|y;\theta) \propto exp\left\{\sum_{i=1}^{k}\left[u_i y_{i+} - \sum_{j=1}^{m_i} log(1 + e^{\beta x_{ij}+u_i}) - \frac{u_i^2}{2\sigma^2}\right]\right\} \qquad (۱.۴)$$

که $y_{i+} = \sum_{j=1}^{m_i} y_{ij}$ برای $i = 1, \ldots, k$ است.

در بخش ۲.۲.۳ ما الگوریتم‌های MHIS، CIS، RQIS و RSIS را معرفی کرده‌ایم. با پذیرفتن شرایط قضیه ۴.۲.۳، این چهار نمونه‌گیری ارگودیک یکنواخت هستند. علاوه بر این، یک نمونه‌گیری متروپلیس قدم زدن تصادفی با بعد کامل، که پیشنهادات پرش به‌طور نرمال توزیع شده‌اند را در نظر می‌گیریم که چگالی پیشنهادی

$$P(u, u^*) \propto exp\left\{\frac{-1}{2\tau^2}\|u^* - u\|_2^2\right\} \qquad (۲.۴)$$

که $\|\cdot\|$ نشان‌دهنده نرم استاندارد اقلیدسی است و τ^2 پارامتر تعدیل ساز است. آنگاه نتیجه زیر برقرار است.

قضیه ۱.۳.۴. *نمونه‌گیری متروپلیس قدم زدن تصادفی با بعد کامل و چگالی مانا رابطه (۱.۴) و چگالی پیشنهادی رابطه (۲.۴) ارگودیک هندسی است.*

برهان. یک نسخه غیرنرمال چگالی هدف در رابطه (۱.۴) به‌دست آمده است. ارگودیک هندسی متروپلیس قدم زدن تصادفی از قضیه ۴.۳ یارنر و هنسن (۲۰۰۰)، پیروی می‌کند اگر ما بتوانیم نشان دهیم

$$\lim_{|u|\to\infty} n(u) \cdot \nabla \log \pi(u) = -\infty$$

و

$$\limsup_{|u|\to\infty} n(u) \cdot m(u) < 0$$

که $n(u)$ نشان‌دهنده بردار واحد $u/|u|$ و $m(u) = \nabla\pi(u)/|\nabla\pi(u)|$ نشان‌دهنده نرم گرادیان π است، به عبارت دیگر

$$\lim_{\|u\|_2\to\infty} \frac{u^T \nabla \log \pi(u)}{\|u\|_2} = -\infty$$

و

$$\limsup_{\|u\|_2\to\infty} \frac{u^T \nabla \log \pi(u)}{\|u\|_2 \|\nabla \log \pi(u)\|_2} < 0$$

برای $i = 1, \ldots, q$ داریم که $\frac{\partial}{\partial u_i} log\pi(u) = y_{i+} - p_{i+} - \frac{u_i}{\sigma^2}$، $p_{i+} = \sum_{m_i}^{j=1} p_{ij}$. بنابراین

$$\begin{aligned}\lim_{\|u\|_2\to\infty} \frac{u^T \nabla \log \pi(u)}{\|u\|_2} &= \lim_{\|u\|_2\to\infty} \frac{\sum_{i=1}^{q} u_i(y_{i+} - p_{i+}) - \frac{1}{\sigma^2}\sum_{i=1}^{q} u_i^2}{\left(\sum_{i=1}^{q} u_i^2\right)^{\frac{1}{2}}} \\ &= -\frac{1}{\sigma^2} \lim_{\|u\|_2\to\infty} \frac{\sum_{i=1}^{q} u_i^2}{\left(\sum_{i=1}^{q} u_i^2\right)^{\frac{1}{2}}} = -\frac{1}{\sigma^2} \lim_{\|u\|_2\to\infty} \|u\|_2 = -\infty\end{aligned}$$

سپس

$$\begin{aligned}\limsup_{\|u\|_۲\to\infty}\frac{u^T\nabla\log\pi(u)}{\|u\|_۲\|\nabla\log\pi(u)\|_۲} &= \limsup_{\|u\|_۲\to\infty}\frac{\sum_{i=۱}^q u_i(y_{i+}-p_{i+})-\frac{۱}{\sigma^۲}\sum_{i=۱}^q u_i^۲}{\left(\sum_{i=۱}^q u_i^۲\right)^{\frac{۱}{۲}}\left(\sum_{i=۱}^q(y_{i+}-p_{i+}-\frac{u_i}{\sigma^۲})^۲\right)^{\frac{۱}{۲}}}\\ &= \limsup_{\|u\|_۲\to\infty}\frac{-\frac{۱}{\sigma^۲}\sum_{i=۱}^q u_i^۲}{\left(\sum_{i=۱}^q u_i^۲\right)^{\frac{۱}{۲}}\left(\frac{۱}{\sigma^۴}\sum_{i=۱}^q u_i^۲\right)^{\frac{۱}{۲}}}\\ &= \frac{\frac{-۱}{\sigma^۲}}{\frac{۱}{\sigma^۲}}\limsup_{\|u\|_۲\to\infty}\frac{\|u\|_۲^۲}{\|u\|_۲^۲}=-۱.\end{aligned}$$

□

پیش از شروع شبیه‌سازی و تحلیل عملکرد تجربی الگوریتم‌ها لازم است که در ابتدا الگوریتم اسکن سیستماتیک و اسکن تصادفی نمونه‌گیری متروپولیس-هستینگز را بیان کنیم.

الگوریتم اسکن سیستماتیک متروپولیس-هستینگز

فرض کنید که حالت اولیه زنجیر، به‌صورت $(X_۱^{(۱)},\ldots,X_۱^{(۲)})$ باشد، زمانی‌که تکرارها $t=۲,۳,\ldots$.

۱. (a) تولید

$$X_۱\sim q_۱(\cdot|X_۱^{(t-۱)},X_۲^{(t-۱)},\ldots,X_d^{(t-۱)})$$

(b) محاسبه

$$\alpha_۱=min\left\{۱,\frac{\pi(X_۱,X_۲^{(t-۱)},\ldots,X_d^{(t-۱)})q(X_۱^{(t-۱)}|X_۱,X_۲^{(t-۱)},\ldots,X_d^{(t-۱)})}{\pi(X_۱^{(t-۱)},X_۲^{(t-۱)},\ldots,X_d^{(t-۱)})q(X_۱|X_۱^{(t-۱)},X_۲^{(t-۱)},\ldots,X_d^{(t-۱)})}\right\}$$

(c) قرار دادن $X_۱^{(t)}=X_۱$ با احتمال $\alpha_۱$ و قرار دادن $X_۱^{(t)}=X_۱^{(t-۱)}$ با احتمال $۱-\alpha_۱$.

...

j. (a) تولید

$$X_j\sim q_j(\cdot|X_۱^{(t)},\ldots,X_{j-۱}^{(t)},X_j^{(t-۱)},\ldots,X_d^{(t-۱)})$$

(b) محاسبه

$$\alpha_j=min\left\{۱,\frac{\pi(X_۱^{(t)},\ldots,X_{j-۱}^{(t)},X_j,X_{j+۱}^{(t-۱)},\ldots,X_d^{(t-۱)})q(X_j^{(t-۱)}|X_۱^{(t)},\ldots,X_{j-۱}^{(t)},X_j,X_{j+۱}^{(t-۱)},\ldots,X_d^{(t-۱)})}{\pi(X_۱^{(t)},\ldots,X_{j-۱}^{(t)},X_j^{(t-۱)},\ldots,X_d^{(t-۱)})q(X_j|X_۱^{(t)},\ldots,X_{j-۱}^{(t)},X_j^{(t-۱)},\ldots,X_d^{(t-۱)})}\right\}$$

(c) قرار دادن $X_j^{(t)}=X_j$ با احتمال α_j و قرار دادن $X_j^{(t)}=X_j^{(t-۱)}$ با احتمال $۱-\alpha_j$.

...

d. (a) تولید

$$X_d\sim q_d(\cdot|X_۱^{(t)},\ldots,X_{d-۱}^{(t)},X_d^{(t-۱)})$$

(b) محاسبه

$$\alpha_d=min\left\{۱,\frac{\pi(X_۱^{(t)},\ldots,X_{d-۱}^{(t)},X_d)q(X_d^{(t-۱)}|X_۱^{(t)},\ldots,X_{d-۱}^{(t)},X_d)}{\pi(X_۱^{(t)},\ldots,X_{d-۱}^{(t)},X_d^{(t-۱)})q(X_d|X_۱^{(t)},\ldots,X_{d-۱}^{(t)},X_d^{(t-۱)})}\right\}$$

(c) قرار دادن $X_d^{(t)}=X_d$ با احتمال α_d و قرار دادن $X_d^{(t)}=X_d^{(t-۱)}$ با احتمال $۱-\alpha_d$.

الگوریتم اسکن تصادفی متروپولیس-هستینگز

فرض کنید که حالت اولیه زنجیر، به‌صورت $(X_۱^{(۱)}, \ldots, X_۱^{(۲)})$ باشد، زمانی‌که تکرارها $t = ۲, ۳, \ldots$.

۱. انتخاب یک شاخص J از توزیعی روی $\{۱, \ldots, d\}$

(a) تولید

$$X_J \sim q_J(\cdot | X_۱^{(t-۱)}, \ldots, X_d^{(t-۱)})$$

(b) محاسبه

$$\alpha_J = min\left\{۱, \frac{\pi(X_۱^{(t-۱)}, \ldots, X_{J-۱}^{(t-۱)}, X_J, X_{J+۱}^{(t-۱)}, \ldots, X_d^{(t-۱)}) q(X_J^{(t-۱)} | X_۱^{(t-۱)}, X_{J-۱}^{(t-۱)}, X_J, X_{J+۱}^{(t-۱)}, \ldots, X_d^{(t-۱)})}{\pi(X_۱^{(t-۱)}, \ldots, X_d^{(t-۱)}) q(X_J | X_۱^{(t-۱)}, \ldots, X_d^{(t-۱)})}\right\}$$

(c) قرار دادن $X_J^{(t)} = X_J$ با احتمال α_J و قرار دادن $X_J^{(t)} = X_J^{(t-۱)}$ با احتمال $۱ - \alpha_J$.

۲. قرار دهید $X_{-J}^{(t)} = X_{-J}^{(t-۱)}$

ما عملکرد تجربی الگوریتم‌ها را در زمینه اجرای الگوریتم مونت کارلو EM ، مقایسه می‌کنیم. هر مرحله MCEM به یک تقریب مونت کارلو که به اصطلاح Q-تابع گفته می‌شود

$$Q(\theta; \tilde{\theta}) = \int \ell_c(\theta; y, u) h(u | y; \tilde{\theta}) du,$$

نیاز دارد به‌طوری‌که

$$\ell_c(\theta; y, u) = \sum_{i=۱}^{k} \sum_{j=۱}^{m_i} [y_{ij}(\beta x_{ij} + u_i) - log(۱ + e^{\beta x_{ij} + u_i})] - \frac{k}{۲} log(\sigma^۲) - \frac{۱}{۲\sigma^۲} \sum_{i=۱}^{k} u_i^۲$$

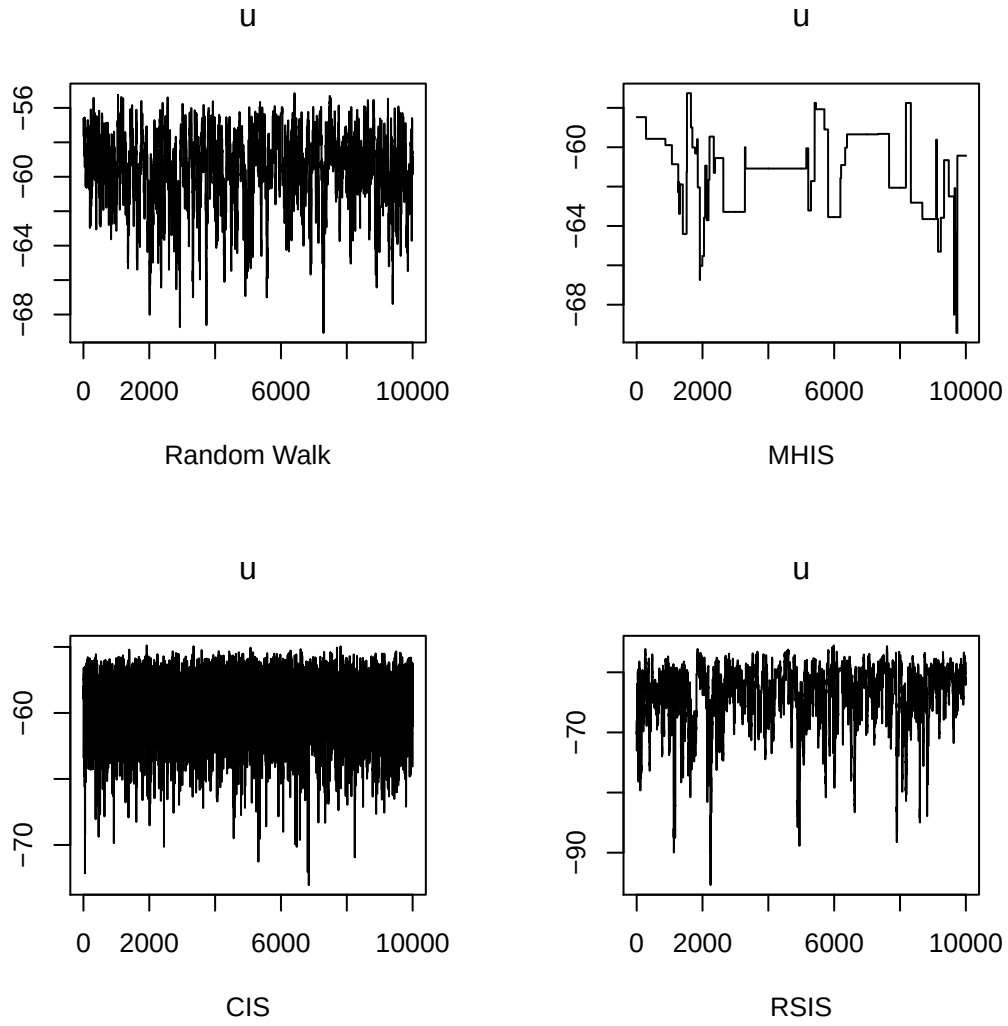
نشان‌دهنده " لگاریتم درستنمایی داده کامل[۱]" است، اگر اثرات تصادفی قابل مشاهده باشد لگاریتم درستنمایی خواهد بود. ما اجرای MCEM را در یک مجموعه داده به عنوان یک مبنا، که توسط بوث و هوبرت (۱۹۹۹) معرفی شده است، در نظر می‌گیریم.

برای اجرای MCEM، فرض می‌کنیم که مقدار واقعی پارامتر $\theta = (\beta, \sigma^۲) = (۵, ۰/۵)$ باشد. در این مجموعه داده، $x_{ij} = \frac{j}{۱۵}$ برای هر $j = ۱, \ldots, m_i = ۱۵$ و برای هر $i = ۱, \ldots, k = ۱۰$ است. فرض کنید $\tilde{\theta} = (۴, ۱/۵)$. می‌توانیم از Q-تابع میانگین نمونه‌ایی زنجیر $\{\ell_c(\theta; y, u^{(t)})\}$، یک تقریب MCEM، بگیریم. به‌طوری‌که

$$Q(\theta; \tilde{\theta}) \approx \bar{\ell}_{C_n}(\theta; \tilde{\theta}) := \frac{۱}{n} \sum_{t=۱}^{n} \ell_c(\theta; y, u^{(t)})$$

$\{u^{(t)} : t = ۱, ۲, \ldots, n\}$ تحققی از پنج زنجیر مارکوف با چگالی مانا $h(u|y; \tilde{\theta})$ است، که به‌وسیله رابطه (۱.۴) تعریف شده است. به‌خاطر سادگی، برآورد نقطه‌ایی $Q(\tilde{\theta}; \tilde{\theta})$ را به‌جای کل تابع در نظر خواهیم گرفت. شرایط آمیختگی در زنجیر مارکوف تضمین کننده وجود CLT برای خطای مونت کارلو $\bar{\ell}_{C_n}(\tilde{\theta}; \tilde{\theta}) - Q(\tilde{\theta}; \tilde{\theta})$، با واریانس توزیع نرمال مجانبی است که با $\sigma_Q^۲$ نشان داده می‌شود که با برآوردگر میانگین دسته‌ای، $\hat{\sigma}_Q^۲$، می‌تواند به برآوردگری سازگار تبدیل شود (جونز و همکاران، ۲۰۰۶).

[۱] Complete- data log-likelihood



شکل ۲.۴: نمودار اثر $\{\ell_c(\theta; y, u^{(t)})\}$ در مدل آمیخته لوجیت نرمال به عنوان مثالی از بخش ۳.۴.

در این مثال الگوریتم‌های MHIS، RW، CIS و RSIS را اجرا کرده‌ایم. در گزارش‌مان از اجرای RQIS فاکتور می‌گیریم، این الگوریتم رفتار مشابهی با CIS در این مثال دارد.

برای شبیه‌سازی این مثال، زنجیری به طول $n = ۱۰^۶$ و با توزیع اولیه $N_{۱۰}(۰, \sigma^۲ I)$ در نظر می‌گیریم. برای متروپلیس قدم‌زدن تصادفی پیشنهادات پرش را از یک $N_{۱۰}(۰, \tau^۲ I)$ با $\tau^۲ = \sigma^۲/۶$ انتخاب می‌کنیم. (این تنظیمات با آزمون و خطا به منظور به حداقل رساندن همبستگی در نتیجه‌گیری زنجیر تعیین شده است).

نمودارهای اثر چهار زنجیر را در شکل (۲.۴) در نظر بگیرید. نتیجه‌ای که نیاز به تامل بیش‌تری دارد این است که با توجه به شکل، MHIS عملکرد بسیار ضعیفی دارد. زنجیر RW سریع‌تر از MHIS آمیخته می‌شود. از طرفی RSIS به نظر می‌رسد که سریع‌تر از MHIS آمیخته می‌شود اما بسیار شبیه به RW رفتار می‌کند. در نهایت، زنجیر CIS به‌تر از این چهار نمونه‌گیری به نظر می‌رسد. می‌توان نتیجه

جدول ۲.۴: نتایج حاصل از مثال بخش ۳.۴، مدل آمیخته لوجیت نرمال.

$Algorithm$	$\hat{Q}(\theta\|\tilde{\theta};y)$	$\hat{\sigma}_Q^2$	$t_*\hat{\sigma}_Q/\sqrt{n}$	ACT	$\widehat{ESEJD}$
RW	$-۵۹/۵۱$	$۰/۰۱۲۵۵$	$۲/۴۵۹۳ \times ۱۰^{-۵}$	$۵/۵۵$	$۰/۵۵۶\,(۰/۰۰۱۲۰۶)$
$MHIS$	$-۵۹/۵۵$	$۰/۰۲۰۹۶$	$۴/۱۰۸ \times ۱۰^{-۵}$	$۹/۵۹$	$۰/۰۷۱۲۹\,(۰/۰۰۱۹)$
CIS	$-۵۹/۵۱$	$۰/۰۰۴۲۳$	$۸/۳۱ \times ۱۰^{-۶}$	$۱/۸۷۵$	$۴/۸۶۳\,(۰/۰۰۵۳)$
$RSIS$	$-۶۳/۷۸$	$۰/۰۲۷۹$	$۵/۴۷۷ \times ۱۰^{-۵}$	$۵/۸۱۶$	$۱/۹۲۱\,(۰/۰۰۶۰۱)$

گرفت زمانی‌که وابستگی ضعیف بین مولفه‌های توزیع هدف وجود داشته باشند، کاربرد CWIS نسبت به MHIS ارجحیت دارد. به‌خاطر بیاورید نیل و رابرتز (۲۰۰۶) که برای متروپلیس قدم زدن تصادفی به یک نتیجه‌گیری مشابه دست یافتند. به‌طور کلی عملکرد تجربی MHIS به دو عامل بستگی دارد: اول، به نزدیکی توزیع پیشنهادی به هدف؛ دوم، به توزیع حاشیه‌ای اثرات تصادفی، U که شباهت کافی با توزیع شرطی U، در داده‌های مفروض ندارد. این نکته قابل ذکر است که با توجه به قضیه ۷.۲.۳ و شکل ۲.۴، MHIS یک زنجیر مارکوف ارگودیک یکنواخت است. بنابراین، این مثال نشان می‌دهد که تکیه بیش از حد به خواص مجانبی نمونه‌گیری، هیچ تضمینی از عملکرد مطلوب در اجرای نمونه‌گیری متناهی فراهم نمی‌کند.

این جدول نتایج حاصل از برآورد $Q(\theta;\tilde{\theta})$ و σ_Q^2 با $\theta = \tilde{\theta} = (۴, ۱/۵)$ براساس $n = ۱۰^۶$ می‌باشد و همچنین ستون میانی جدول نشان‌دهنده نصف پهنای فاصله اطمینان ۹۵ درصدی با $t_* = ۱/۹۶۰$ است و ACT نیز بیانگر معیار همبستگی یکپارچه زمان می‌باشد.

نتایج شبیه‌سازی ارایه شده در جدول (۲.۴) را در نظر بگیرید. با به‌کارگیری اندازه نمونه مونت کارلو یکسان، نصف پهنای برآوردگر فاصله‌ای ارایه شده، با کمی اغماض، برای کلیه الگوریتم‌ها تقریبا یکسان است. البته نصف پهنا برای CIS از بقیه کم‌تر است پس می‌توان گفت که همگرایی در این الگوریتم سریع‌تر صورت می‌گیرد.

با توجه به معیار ACT، الگوریتم CIS دارای کم‌ترین مقدار نسبت به سایر الگوریتم‌هاست. بر اساس این معیار نیز، کاراتر بودن این الگوریتم تایید می‌شود. MHIS دارای بیش‌ترین ACT است بنابراین عملکرد مناسبی نسبت به بقیه ندارد. برآورد ESEJD بر اساس همان $m = ۵۰۰$ تکرار مستقل از RW، MHIS، RSIS و CIS است. با اختلاف و اطمینان بیش‌تری می‌توان گفت CIS عملکرد به‌تری دارد اما MHIS به دلیل کوچک بودن مقدار ESEJD دارای عملکرد پایین است. همچنین می‌توان گفت که RSIS و RW رفتار یکسانی دارند.

جای تعجب دارد که در اجرای RSIS در یک مرحله، تنها یکی از ده مولفه‌ها شانس به‌روز شدن دارد، اما عملکرد آن شبیه زنجیری است که همه مولفه‌هایش حدود ۳۰ درصد از زمان به‌روز رسانی می‌شوند.

پیشنهادات و آینده تحقیق

به غیر از نمونه‌گیر گیبس دو متغیره و اسکن تصادفی الگوریتم متروپولیس–هستینگز درون گیبز، تحقیقات کمی روی نرخ همگرایی، نمونه‌گیری‌های MCMC مولفه به مولفه به‌ویژه MH صورت گرفته است. ارگودیک هندسی یک ویژگی بسیار مطلوب برای یک زنجیر مارکوف است، به این دلیل که در اکثر موارد MH برگشت‌پذیر است (گیر، ۲۰۱۱)، MH همیشه یک زنجیر مارکوف برگشت‌پذیر با توزیع مانای $\pi(\cdot)$ صرف‌نظر از این‌که چگالی پیشنهادی چگونه انتخاب می‌شود، تولید می‌کند. ایجاد ارگودیک هندسی یک گام کلیدی است زیرا برای کاربران آن‌ها این اطمینان را حاصل می‌کند که می‌توانند از نمونه‌های تولید شده از چنین زنجیرهایی با عنوان نمونه‌های با کیفیت نزدیک به نمونه‌های یکسان، همتوزیع و مستقل استفاده کنند.

روش های MCMC مولفه به مولفه می توانند نسبت به به‌روز رسانی با بعد کامل برتری داشته باشند، برای مثال روش‌های MCMC با بعد کامل، اغلب برای این‌که ارگودیک هندسی باشند با شکست روبه‌رو می‌شوند. همچنین بررسی‌های تجربی فصل ۴ نشان داد که ویژگی‌های نمونه متناهی از روش‌های مولفه به مولفه نسبت به روش‌های با بعد کامل در هر حالتی برتری دارد، که ما می‌توانیم مشاهدات‌مان را برای بسیاری از داده‌های واقعی تطبیق دهیم.

یکی از مباحثی که در این پایان‌نامه به آن پرداختیم، مطالعه نرخ همگرایی نمونه‌های مولفه به مولفه تشکیل شده به‌وسیله ترکیب، P_C ما را قادر به ایجاد ارگودیک هندسی و یکنواخت برای دیگر مولفه‌ها مانند P_{RQ} و P_{RS} می‌سازد. ما نشان دادیم که این برای ارگودیک یکنواخت در تنظیمات کلی و برای ارگودیک هندسی در تنظیمات دو متغیره درست است. به نظر می‌رسد که مطالعه نرخ همگرایی P_{RQ} و P_{RS} به ما درباره نرخ P_C اطلاع می‌دهد.

پیوست الف

در این پیوست، متن برنامه‌هایی که در محیط نرم افزار R برای شبیه‌سازی فصل چهارم نوشته شده، گزارش شده است. شبیه‌سازی اول مربوط به مدل آمیخته خطی بیزی و شبیه‌سازی دوم برای مدل آمیخته لجستیک نرمال می‌باشد.

- بسته‌های فراخوانی شده برای شبیه‌سازی مدل اول

```
rm(list=ls())
library(MASS)
library(base)
library(magic)
library(MCMCglmm)
library(mvtnorm)
library(mcmc)
set.seed(3241)
```

از مدل آمیخته خطی بیزی تولید نمونه می‌کنیم

```
gen.data = function(k=10, m=5, p=1){
N = k*m
B = 0.1
r1 = r2 = d1 = d2 = 2
xx = c(-0.5,-0.25,0,0.25,0.5)
x = rep(xx,k)
X = matrix(x,nrow=N)
beta = rnorm(1,0,B)
```

```
I = diag(10)
l= rep(1, m)
lambda.d = rgamma(1, shape=d1, rate=d2)
u = mvrnorm(1, rep(0, 10), 1/(lambda.d)*diag(10))
Z = kronecker(I,l)
M = X%*%beta+Z%*%u
lambda.r = rgamma(1, shape=r1, rate=r2)
S = (1/lambda.r)*diag(N)
y = mvrnorm(1,M,S)
return(list(y = y, X = X, Z = Z, x = x, u = u, beta = beta))
}

data = gen.data(k=10, m=5, p=1)

y = data$y; X = data$X; Z = data$Z; x = data$x; u = data$u
k = 10
m = 5
p = 1
N = k*m

####
#### Gibbs sampling

GS.MCMC = function(Nsim, B=0.1, b = 0){
B = 1/B
r1 = r2 = d1 = d2 = 3
param = matrix(0, nrow=Nsim, ncol=13)
 param[1,] = c(0, rep(0, 10), 1, 1)
      for (i in 2:Nsim){
       A1 = diag(1/diag(param[i-1,12]*t(Z)%*%Z+param[i-1,13]*diag(k)))
          A2 = t(Z)%*%y
          B1 = 1/(param[i-1,12]*t(X)%*%X+B)
          B2 = (param[i-1,12]*t(x)%*%y)+B*b
          comp1 = param[i-1,12]*(A1%*%A2)
```

```
          comp2 = B1*B2
          m0 <- c(comp1,comp2)
          Sigma.inv = adiag(A1,B1)
 param[i,-(12:13)] = mvrnorm(1,m0,Sigma.inv)
 m = X%*%param[i,1]+Z%*%param[i,2:11]
 nu1 <- t(y-m)%*%(y-m)
 nu2<-t(param[i,2:11])%*%param[i,2:11]
      param[i,12] = rgamma(1, shape=r1+(N/2), rate=r2+(nu1/2))
      param[i,13] = rgamma(1, shape=d1+(k/2), rate=d2+(nu2/2))
 }
 return(list(u=param[,2:11], beta=param[,1], lam.r=param[,12],
lam.d=param[,13]))
 }

 result.GS = GS.MCMC(Nsim = 100000)

####
#### randam scan Gibbs sampling

RS.MCMC = function(Nsim, B=0.1, b=0){
B = 1/B
r1 = r2 = d1 = d2 = 3
param = matrix(0, nrow=Nsim, ncol=13)
param[1,] = c(0, rep(0, 10), 1, 1)
c1 = c2 = 1/2
for (i in 2:Nsim){
w = sample(1:2, size=1, prob=c(c1,c2))
if(w==1){
A1 = solve(param[i-1,12]*t(Z)%*%Z+param[i-1,13]*diag(k))
A2 = t(Z)%*%y
B1 = 1/(param[i-1,12]*t(X)%*%X+B)
B2 = param[i-1,12]*t(x)%*%y+B*b
comp1 = param[i-1,12]*(A1%*%A2)
comp2 = B1*B2
```

```
m0 <- c(comp1,comp2)
Sigma.inv = adiag(A1,B1)

param[i,-(12:13)] = mvrnorm(1,m0,Sigma.inv)
param[i,12] = param[i-1,12]
param[i,13] = param[i-1,13]
}
if(w==2){
m = X%*%param[i-1,1]+Z%*%param[i-1,2:11]
nu1 <- t(y-m)%*%(y-m)
nu2<-t(param[i-1,2:11])%*%param[i-1,2:11]
param[i,12] = rgamma(1, shape=r1+(N/2), rate=r2+(nu1/2))
param[i,13] = rgamma(1, shape=d1+(k/2), rate=d2+(nu2/2))
param[i,-(12:13)] = param[i-1,-(12:13)]
}
}
return(list(beta=param[,1], u=param[,2:11],
 lam.r=param[,12], lam.d=param[,13]))
}

result.RS = RS.MCMC(Nsim = 100000)

####
#### random sequence Gibbs sampling

RSeq.MCMC = function(Nsim, B = 0.1, b = 0){
B = 1/B
r1 = r2 = d1 = d2 = 3
param = matrix(0, nrow=Nsim, ncol=13)
param[1,] = c(0, rep(0, 10), 1, 1)
for (i in 2:Nsim){
G = sample(1:2, size=1, prob=c(1/2,1/2))
if (G==1) {
A1 = diag(1/diag(param[i-1,12]*t(Z)%*%Z+param[i-1,13]*diag(k)))
```

```
A2 = t(Z)%*%y
B1 = 1/(param[i-1,12]*t(X)%*%X+B)
B2 = param[i-1,12]*t(x)%*%y+B*b
B3 = solve(param[i-1,12]*t(x)%*%y+B)
comp1 = param[i-1,12]*(A1%*%A2)
comp2 = B1*B2
m0 <- c(comp1,comp2)
Sigma.inv = adiag(A1,B2)

param[i,-(12:13)] = mvrnorm(1,m0,Sigma.inv)

m = X%*%param[i,1]+Z%*%param[i,2:11]
nu1 <- t(y-m)%*%(y-m)
nu2<-t(param[i,2:11])%*%param[i,2:11]
param[i,12] = rgamma(1, shape=r1+(N/2), rate=r2+(nu1/2))
param[i,13] = rgamma(1, shape=d1+(k/2), rate=d2+(nu2/2))
}
else {

m = X%*%param[i-1,1]+Z%*%param[i-1,2:11]
nu1 <- t(y-m)%*%(y-m)
nu2<-t(param[i-1,2:11])%*%param[i-1,2:11]
param[i,12] = rgamma(1, shape=r1+(N/2), rate=r2+(nu1/2))
param[i,13] = rgamma(1, shape=d1+(k/2), rate=d2+(nu2/2))

A1 = diag(1/diag(param[i-1,12]*t(Z)%*%Z+param[i-1,13]*diag(k)))
A2 = t(Z)%*%y
B1 = 1/(param[i,12]*t(X)%*%X+B)
B2 = param[i,12]*t(x)%*%y+B*b
B3 = param[i,12]*t(x)%*%y+B
comp1 = param[i,12]*(A1%*%A2)
comp2 = B1*B2
m0 <- c(comp1,comp2)
Sigma.inv = adiag(A1,B3)
```

```
param[i,-(12:13)] = mvrnorm(1,m0,Sigma.inv)
}
}
return(list(beta=param[,1], lam.r=param[,12], lam.d=param[,13]))
}

result.RSeq = RSeq.MCMC(Nsim = 100000)

####
#### MH-RW MCMC Run

%set the diagonal of covariance matrix of proposal for (u,beta)
Y = cbind(result.GS$beta, result.GS$u, result.GS$lam.r, result.GS$lam.d)
prop.s = diag(diag(cov(Y)^2))

B = 1/0.1
b = 0
r1=r2=d1=d2=3

log.posterior = function(param, k=10){
beta = param[1]
u = param[2:11]
lambda.r = param[12]
lambda.d = param[13]
m = X%*%beta+Z%*%u

log.post = dmvnorm(y, mean = m, sig = (1/lambda.r)*diag(N),
log=TRUE)+  dnorm(beta, mean = b, sd = B, log=TRUE)+
dmvnorm(u, mean = rep(0,k), sig = (1/lambda.d)*diag(k),
log=TRUE) + dgamma(lambda.r, shape = r1, rate = r2, log=TRUE) +
dgamma(lambda.d, shape = d1, rate = d2, log=TRUE)
return(log.post)
}
```

```
Nsim = 100000
param = matrix(0,ncol=13, nrow = Nsim)

RWM.MCMC<-function(Nsim){
k = 10
      b=0
      B=1/(0.1)
      param[1,1:11] = rep(0,11)
      param[1,12:13]=rep(1,2)
      k=0
      for(i in 2:Nsim){
    prop.xi = param[i-1,] + as.vector(rmvnorm(1,
mean = rep(0,13), sigma = prop.s))

    R1 = log.posterior(param = prop.xi)
    R2 = log.posterior(param = param[i-1,])
    ratio = exp(R1-R2)
    rho=min(1,ratio, na.rm = TRUE)
          if(runif(1)< rho){
          param[i,]=prop.xi
    k = k+1 }
          else{
     param[i,] = param[i-1,]
               }
         }
      k = k/Nsim

      list(beta=param[,1], lam.r=param[,12], lam.d=param[,13], k=k)
}


result.rw = RWM.MCMC(Nsim = 100000)

beta.star = mean(result.rw$beta)
```

```
par(mfrow=c(2,2))
ts.plot(result.GS$beta[99000:100000], xlab="Gibbs Sampling",
 main = expression(beta),
ylab = "")
ts.plot(result.RS$beta[99000:100000], xlab="Random scan",
 main = expression(beta),
ylab = "")
ts.plot(result.RSeq$beta[99000:100000], xlab="Random sequence",
main = expression(beta),
ylab = "")
ts.plot(result.rw$beta[99000:100000], xlab="Random walk MH",
 main = expression(beta),
ylab = "")
```

سپس با ۵۰۰ تکرار، زنجیر مارکوف مونت‌کارلو را شبیه‌سازی نموده و مقادیر ، $\bar{\beta}_n$، $\hat{\sigma}^2_\beta$، $t_* \hat{\sigma}_\beta / \sqrt{n}$، ACT، $MSE ratio$ و $\widehat{ESEJD}$ را به‌دست می‌آوریم.

```
iter = 500

## column 1

bta.bar.GS = mean(result.GS$beta)
bta.bar.RS = mean(result.RS$beta)
bta.bar.RSeq = mean(result.RSeq$beta)
bta.bar.rw = mean(result.rw$beta)

## column 2

sigma.beta.GS = var(result.GS$beta)/Nsim
sigma.beta.RS = var(result.RS$beta)/Nsim
sigma.beta.RSeq = var(result.RSeq$beta)/Nsim
sigma.beta.rw = var(result.rw$beta)/Nsim

### column 3
```

```
t.star = 1.96

inter.GS  <- t.star*sqrt(sigma.beta.GS/Nsim)
inter.RS <- t.star*sqrt(sigma.beta.RS /Nsim)
inter.RSeq <- t.star*sqrt(sigma.beta.RSeq /Nsim)
inter.rw <- t.star*sqrt(sigma.beta.rw /Nsim)

## column 4

xx.GS = result.GS$beta
nn.GS = length(xx.GS)
batch.GS = mcse(xx.GS, size=100)$se
ACT.GS = batch.GS/sqrt(var(xx.GS)/nn.GS)


xx.RS = result.RS$beta
nn.RS = length(xx.RS)
batch.RS = mcse(xx.RS, size=100)$se
ACT.RS = batch.RS/sqrt(var(xx.RS)/nn.RS)


xx.RSeq = result.RSeq$beta
nn.RSeq = length(xx.RSeq)
batch.RSeq = mcse(xx.RSeq, size=100)$se
ACT.RSeq = batch.RSeq/sqrt(var(xx.RSeq)/nn.RSeq)


xx.rw = result.rw$beta
nn.rw = length(xx.rw)
batch.rw = mcse(xx.rw, size=100)$se
ACT.rw = batch.rw/sqrt(var(xx.rw)/nn.rw)

## column 5,6
```

```
MSE.GS<- function(Nsim = 10000, iter ){
result.GS = result.RS = result.RSeq = result.rw = rep(0,iter)
JD.GS<-JD.RS<-JD.RSeq<-JD.rw <- rep(0,iter)
for (i in 1:iter){
data = gen.data(k=10, m=5, p=1)
y = data$y; X = data$X; Z = data$Z; x = data$x; u = data$u
N = k*m

GS.beta = GS.MCMC(Nsim)$beta
RS.beta = RS.MCMC(Nsim)$beta
RSeq.beta = RSeq.MCMC(Nsim)$beta
rw.beta = RWM.MCMC(Nsim)$beta

result.GS[i] = mean(GS.beta)
result.RS[i] = mean(RS.beta)
result.RSeq[i] = mean(RSeq.beta)
result.rw[i] = mean(rw.beta)

# Estimating MSEJD

JD.GS[i] = mean(diff(GS.beta)^2)
JD.RS[i] = mean(diff(RS.beta)^2)
JD.RSeq[i] =mean(diff(RSeq.beta)^2)
JD.rw[i] = mean(diff(rw.beta)^2)
print(c(i,date()))
}

#MSE

      MSE.GS = mean((result.GS-beta.star)^2)
      MSE.RS = mean((result.RS-beta.star)^2)
      MSE.RSeq = mean((result.RSeq-beta.star)^2)
      MSE.rw = mean((result.rw-beta.star)^2)
```

```
#Ratio

      R.MSE.GS = MSE.GS/MSE.GS
      R.MSE.RS = MSE.RS/MSE.GS
      R.MSE.RSeq = MSE.RSeq/MSE.GS
      R.MSE.rw = MSE.rw/MSE.GS

#Estimating standard errors of MSE ratio estimates

     MSE.se.GS =  sd((result.GS-beta.star)^2/(result.GS-beta.star)^2)
    MSE.se.RS  = sd( (result.RS-beta.star)^2/(result.GS-beta.star)^2)
   MSE.se.RSeq = sd( (result.RSeq-beta.star)^2/(result.GS-beta.star)^2)
     MSE.se.rw  = sd((result.rw-beta.star)^2/(result.GS-beta.star)^2)

 # Estimating ESEJD

    ESEJD.GS = mean(JD.GS)
ESEJD.RS = mean(JD.RS)
ESEJD.RSeq = mean(JD.RSeq)
ESEJD.rw = mean(JD.rw)

# Estimating standard errors of ESEJD estimates

    ESEJD.se.GS = sd(JD.GS)/sqrt(iter)
ESEJD.se.RS = sd(JD.RS)/sqrt(iter)
ESEJD.se.RSeq = sd(JD.RSeq)/sqrt(iter)
ESEJD.se.rw = sd(JD.rw)/sqrt(iter)


list(R.MSE.GS = R.MSE.GS, R.MSE.RS=R.MSE.RS, R.MSE.RSeq=R.MSE.RSeq,
      R.MSE.rw=R.MSE.rw,
    MSE.se.GS = MSE.se.GS, MSE.se.RS = MSE.se.RS, MSE.se.RSeq = MSE.se.RSeq,
      MSE.se.rw = MSE.se.rw,
```

```
ESEJD.GS = ESEJD.GS, ESEJD.RS = ESEJD.RS, ESEJD.RSeq =ESEJD.RSeq,
      ESEJD.rw = ESEJD.rw,
    ESEJD.se.GS =ESEJD.se.GS, ESEJD.se.RS = ESEJD.se.RS ,ESEJD.se.RSeq = ESEJD.se.R
      ESEJD.se.rw =ESEJD.se.rw )
}

output = MSE.GS(Nsim = 10000, iter=2)
output
```

- بسته‌های فراخوانی شده برای شبیه‌سازی مدل دوم

```
rm(list=ls())
library(MASS)
library(base)
library(magic)
library(MCMCglmm)
library(mvtnorm)
set.seed(12435)
```

از مدل آمیخته لجستیک نرمال تولید نمونه می‌کنیم

```
gen.data = function(k=10, m=15, sig2=0.5, beta=5){
N = k*m
      X =  matrix(,ncol = m, nrow = 10)
      for (j in 1:m){
             X[,j] = rep(j/m,10)
                    }
u = rnorm(k, 0, sig2)
eta = beta*X+u
      p = exp(eta) / (1 + exp(eta))
y = matrix(rbinom(N,1,p), ncol = m, byrow = T)
return(list(y = y, X = X, u = u))
}

data = gen.data()
```

```
y = data$y; X = data$X; u = data$u
k = 10
m = 15
N = k*m

####
####target distribution

target.den = function(y, X, u, beta, sig2, k, m){
yi. = apply(y, 1, sum)
a1 = (beta*X)+u
a2 = u*yi.
a3 = apply(log(1+exp(a1)), 1, sum)
a4 = u^2/(2*sig2)
h.u = exp(sum(a2-a3-a4))
return(h.u)
}

# log likelihood complement

L.L.C = function(y, X, u, beta, sig2, k){
a = (beta*X)+u
a1 = y*a
a2 = log(1+exp(a))
a3 = sum(a1 - a2)
a4 = (k/2)*log(sig2)
a5 = (1/(2*sig2))*sum(u^2)
L.c = a3 - a4 - a5
return(L.c)
}
####
#### Random Walk Algorithm
```

```
RWM.MCMC<-function(Nsim){
      beta <- 4
      sig2 <- 1.5
k <- 10
      m <- 15
     u <- matrix( ,ncol = k, nrow = Nsim)          # Empty contain for u
     u[1, ] <- rmvnorm(1, rep(0, k), sig2*diag(k))   # First value for u
      a.r <- 0                                       # Acceptance rate
      for(i in 2:Nsim){
          prop.den <- as.vector(rmvnorm(1, u[i-1,], (sig2/6)*diag(k)))
           ratio <- target.den(y, X, prop.den, beta, sig2, k, m)/
                    target.den(y, X, u[i-1,], beta, sig2, k, m)
    rho <- min(1,ratio, na.rm = TRUE)
          if(runif(1)< rho){
          u[i,] <- prop.den
    a.r = a.r+1 }
          else{
     u[i,] <- u[i-1,]
               }
         }
      a.r <- a.r/Nsim
# output
      list(u=u, a.r=a.r)
}

Nsim=10^6
result.RW <- RWM.MCMC(Nsim=Nsim)

result.RW$a.r

lik.comp.RW <- rep(0,Nsim)
for (i in 1:Nsim){
lik.comp.RW[i] <- L.L.C(y,X,result.RW$u[i,],beta=4, sig2=1.5, k=10)
}
```

```
####
#### MHIS Algorithm

MHIS.MCMC <-function(Nsim){
      beta <- 4
sig2 <- 1.5
      k <- 10
      m <- 15
     u <- matrix( ,ncol = k, nrow = Nsim)          # Empty contain for u
     u[1, ] <- rmvnorm(1, rep(0, k), sig2*diag(k))   # First value for u
      a.r <- 0                                       # Acceptance rate
      for(i in 2:Nsim){
          prop.den <- as.vector(rmvnorm(1, rep(0, k), sig2*diag(k)))
          S <- log(target.den(y, X, prop.den, beta, sig2, k, m))+
               dmvnorm(u[i-1, ], rep(0, k), sig2*diag(k), log=TRUE)
          M <- log(target.den(y, X, u[i-1, ], beta, sig2, k, m))+
               dmvnorm(prop.den, rep(0, k), sig2*diag(k), log=TRUE)
          ratio <- exp(S-M)
    rho <- min(1,ratio, na.rm = TRUE)
          if(runif(1)< rho){
          u[i,] <- prop.den
    a.r = a.r+1 }
          else{
     u[i,] <- u[i-1,]
               }
         }
      a.r <- a.r/Nsim

      list(u=u, a.r=a.r)
}

result.MHIS <- MHIS.MCMC(Nsim=Nsim)
result.MHIS$a.r
```

```
lik.comp.MHIS <- rep(0,Nsim)
for (i in 1:Nsim){
lik.comp.MHIS[i] <- L.L.C(y,X,result.MHIS$u[i,],beta=4, sig2=1.5, k=10)
}

####
#### CIS Algorithm

CIS.MCMC <- function(Nsim){
      beta <- 4
      sig2 <- 1.5
k <- 10
      m <- 15
      tau <- sig2/6
     u <- matrix( ,ncol = k, nrow = Nsim)          # Empty contain for u
     u[1, ] <- rmvnorm(1, rep(0, k), sig2*diag(k))   # First value for u
      a.r <- 0                                       # Acceptance rate
      for(i in 2:Nsim){
### updating the first block u[,1]
p1 <- rnorm(1,0,sig2)
prop.den1 <- c(p1,u[i-1,2:k])
           S1 <- log(target.den(y, X, prop.den1, beta, sig2, k, m))+
dnorm(u[i-1,1],0,sig2,log=TRUE)
           M1 <- log(target.den(y, X, u[i-1, ], beta, sig2, k, m))+
                dnorm(p1,0,sig2,log=TRUE)
           ratio1 <- exp(S1-M1)
     rho1 <- min(1,ratio1, na.rm = TRUE)
           if(runif(1)< rho1){
           u[i,1] <- p1
}
else{
     u[i,1] <- u[i-1,1]
            }
### updating the second block u[,2]
```

```
p2 <- rnorm(1,0,sig2)
prop.den2 <- c(u[i,1],p2,u[i-1,3:k])
          S2 <- log(target.den(y, X, prop.den2, beta, sig2, k, m))+
dnorm(u[i-1,2],0,sig2,log=TRUE)
       M2 <- log(target.den(y, X, c(u[i,1],u[i-1,2:k]), beta, sig2, k, m))+
               dnorm(p2,0,sig2,log=TRUE)
          ratio2 <- exp(S2-M2)
     rho2 <- min(1,ratio2, na.rm = TRUE)
          if(runif(1)< rho2){
          u[i,2] <- p2
}
else{
     u[i,2] <- u[i-1,2]
           }
### updating the third block u[,3]
p3 <- rnorm(1,0,sig2)
prop.den3 <- c(u[i,1],u[i,2],p3,u[i-1,4:k])
          S3 <- log(target.den(y, X, prop.den3, beta, sig2, k, m))+
dnorm(u[i-1,3],0,sig2,log=TRUE)
       M3 <- log(target.den(y, X, c(u[i,1:2],u[i-1,3:k]), beta, sig2, k, m))+
               dnorm(p3,0,sig2,log=TRUE)
          ratio3 <- exp(S3-M3)
     rho3 <- min(1,ratio3, na.rm = TRUE)
          if(runif(1)< rho3){
          u[i,3] <- p3
}
else{
     u[i,3] <- u[i-1,3]
           }
### updating the 4th block u[,4]
p4 <- rnorm(1,0,sig2)
prop.den4 <- c(u[i,1:3],p4,u[i-1,5:k])
          S4 <- log(target.den(y, X, prop.den4, beta, sig2, k, m))+
dnorm(u[i-1,4],0,sig2,log=TRUE)
```

```
       M4 <- log(target.den(y, X, c(u[i,1:3],u[i-1,4:k]), beta, sig2, k, m))+
                dnorm(p4,0,sig2,log=TRUE)
           ratio4 <- exp(S4-M4)
     rho4 <- min(1,ratio4, na.rm = TRUE)
           if(runif(1)< rho4){
           u[i,4] <- p4
}
else{
     u[i,4] <- u[i-1,4]
            }
### updating the 5th block u[,5]
p5 <- rnorm(1,0,sig2)
prop.den5 <- c(u[i,1:4],p5,u[i-1,6:k])
           S5 <- log(target.den(y, X, prop.den5, beta, sig2, k, m))+
dnorm(u[i-1,5],0,sig2,log=TRUE)
       M5 <- log(target.den(y, X, c(u[i,1:4],u[i-1,5:k]), beta, sig2, k, m))+
                dnorm(p5,0,sig2,log=TRUE)
           ratio5 <- exp(S5-M5)
     rho5 <- min(1,ratio5, na.rm = TRUE)
           if(runif(1)< rho5){
           u[i,5] <- p5
}
else{
     u[i,5] <- u[i-1,5]
            }
### updating the 6th block u[,6]
p6 <- rnorm(1,0,sig2)
prop.den6 <- c(u[i,1:5],p6,u[i-1,7:k])
           S6 <- log(target.den(y, X, prop.den6, beta, sig2, k, m))+
dnorm(u[i-1,6],0,sig2,log=TRUE)
       M6 <- log(target.den(y, X, c(u[i,1:5],u[i-1,6:k]), beta, sig2, k, m))+
                dnorm(p6,0,sig2,log=TRUE)
           ratio6 <- exp(S6-M6)
     rho6 <- min(1,ratio6, na.rm = TRUE)
```

```
           if(runif(1)< rho6){
           u[i,6] <- p6
}
else{
     u[i,6] <- u[i-1,6]
            }
### updating the 7th block u[,7]
p7 <- rnorm(1,0,sig2)
prop.den7 <- c(u[i,1:6],p7,u[i-1,8:k])
           S7 <- log(target.den(y, X, prop.den7, beta, sig2, k, m))+
dnorm(u[i-1,7],0,sig2,log=TRUE)
       M7 <- log(target.den(y, X, c(u[i,1:6],u[i-1,7:k]), beta, sig2, k, m))+
                dnorm(p7,0,sig2,log=TRUE)
           ratio7 <- exp(S7-M7)
     rho7 <- min(1,ratio7, na.rm = TRUE)
           if(runif(1)< rho7){
           u[i,7] <- p7
}
else{
     u[i,7] <- u[i-1,7]
            }
### updating the 8th block u[,8]
p8 <- rnorm(1,0,sig2)
prop.den8 <- c(u[i,1:7],p8,u[i-1,9:k])
           S8 <- log(target.den(y, X, prop.den8, beta, sig2, k, m))+
dnorm(u[i-1,8],0,sig2,log=TRUE)
       M8 <- log(target.den(y, X, c(u[i,1:7],u[i-1,8:k]), beta, sig2, k, m))+
                dnorm(p8,0,sig2,log=TRUE)
           ratio8 <- exp(S8-M8)
     rho8 <- min(1,ratio8, na.rm = TRUE)
           if(runif(1)< rho8){
           u[i,8] <- p8
}
else{
```

```
     u[i,8] <- u[i-1,8]
            }
### updating the 9th block u[,9]
p9 <- rnorm(1,0,sig2)
prop.den9 <- c(u[i,1:8],p9,u[i-1,k])
          S9 <- log(target.den(y, X, prop.den9, beta, sig2, k, m))+
dnorm(u[i-1,9],0,sig2,log=TRUE)
       M9 <- log(target.den(y, X, c(u[i,1:8],u[i-1,9:k]), beta, sig2, k, m))+
               dnorm(p9,0,sig2,log=TRUE)
          ratio9 <- exp(S9-M9)
     rho9 <- min(1,ratio9, na.rm = TRUE)
          if(runif(1)< rho9){
          u[i,9] <- p9
}
else{
     u[i,9] <- u[i-1,9]
            }
### updating the 10th block u[,10]
p10 <- rnorm(1,0,sig2)
prop.den10 <- c(u[i,1:9],p10)
         S10 <- log(target.den(y, X, prop.den10, beta, sig2, k, m))+
dnorm(u[i-1,10],0,sig2,log=TRUE)
       M10 <- log(target.den(y, X, c(u[i,1:9],u[i-1,k]), beta, sig2, k, m))+
               dnorm(p10,0,sig2,log=TRUE)
          ratio10 <- exp(S10-M10)
     rho10 <- min(1,ratio10, na.rm = TRUE)
          if(runif(1)< rho10){
          u[i,10] <- p10
}
else{
     u[i,10] <- u[i-1,10]
            }
}
```

```
      list(u=u)
}

result.CIS <- CIS.MCMC(Nsim=Nsim)
result.CIS$a.r

lik.comp.CIS <- rep(0,Nsim)
for (i in 1:Nsim){
lik.comp.CIS[i] <- L.L.C(y,X,result.CIS$u[i,],beta=4, sig2=1.5, k=10)
}

####
#### RSIS Algorithm

RSIS.MCMC <- function(Nsim){
      p1 <- p2 <- p3 <- p4 <- p5 <- p6 <- p7 <- p8 <- p9 <- p10 <- 0
beta <- 4
      sig2 <- 1.5
k <- 10
      m <- 15
      tau <- sig2/6
     u <- matrix( ,ncol = k, nrow = Nsim)          # Empty contain for u
     u[1, ] <- rmvnorm(1, rep(0, k), sig2*diag(k))   # First value for u
      a.r <- 0
      for (i in 2:Nsim){
      w = sample(1:10, size=1, prob=rep(0.1,10))
if(w==1){
          p1 <- rnorm(1,0,sig2)
          prop.den <- c(p1,u[i-1,2:10])
}
if(w==2){
          p2 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1],p2,u[i-1,3:10])
}
```

```
if(w==3){
          p3 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:2],p3,u[i-1,4:10])
}
if(w==4){
          p4 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:3],p4,u[i-1,5:10])
}
if(w==5){
          p5 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:4],p5,u[i-1,6:10])
}
if(w==6){
          p6 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:5],p6,u[i-1,7:10])
}
if(w==7){
          p7 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:6],p7,u[i-1,8:10])
}
if(w==8){
          p8 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:7],p8,u[i-1,9:10])
}
if(w==9){
          p9 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:8],p9,u[i-1,10])
}
if(w==10){
          p10 <- rnorm(1,0,sig2)
          prop.den <- c(u[i-1,1:9],p10)
}
          S <- log(target.den(y, X, prop.den, beta, sig2, k, m))+
                dnorm(u[i-1,1],0,sig2,log=TRUE)+
```

```
dnorm(u[i-1,2],0,sig2,log=TRUE)+
dnorm(u[i-1,3],0,sig2,log=TRUE)+
                  dnorm(u[i-1,4],0,sig2,log=TRUE)+
dnorm(u[i-1,5],0,sig2,log=TRUE)+
dnorm(u[i-1,6],0,sig2,log=TRUE)+
                  dnorm(u[i-1,7],0,sig2,log=TRUE)+
dnorm(u[i-1,8],0,sig2,log=TRUE)+
dnorm(u[i-1,9],0,sig2,log=TRUE)+
                dnorm(u[i-1,10],0,sig2,log=TRUE)
          M <- log(target.den(y, X, u[i-1, ], beta, sig2, k, m))+
                dnorm(p1,0,sig2,log=TRUE)+
dnorm(p2,0,sig2,log=TRUE)+
dnorm(p3,0,sig2,log=TRUE)+
                dnorm(p4,0,sig2,log=TRUE)+
dnorm(p5,0,sig2,log=TRUE)+
dnorm(p6,0,sig2,log=TRUE)+
                  dnorm(p7,0,sig2,log=TRUE)+
dnorm(p8,0,sig2,log=TRUE)+
dnorm(p9,0,sig2,log=TRUE)+
            dnorm(p10,0,sig2,log=TRUE)
          ratio <- exp(S-M)
    rho <- min(1,ratio, na.rm = TRUE)
          if(runif(1)< rho){
          u[i,] <- prop.den
    a.r = a.r+1 }
          else{
     u[i,] <- u[i-1,]
              }
         } #end of loop (for)
       a.r <- a.r/Nsim

       list(u=u, a.r=a.r)
}
```

```
result.RSIS <- RSIS.MCMC(Nsim=Nsim)
result.RSIS$a.r

lik.comp.RSIS <- rep(0,Nsim)
for (i in 1:Nsim){
lik.comp.RSIS[i] <- L.L.C(y,X,result.RSIS$u[i,],beta=4, sig2=1.5, k=10)
}

#ts.plot
par(mfrow=c(2,2))

ts.plot(lik.comp.RW[10000:20000], xlab="Random Walk", main = expression(u),
ylab = "")

ts.plot(lik.comp.MHIS[10000:20000], xlab="MHIS", main = expression(u),
ylab = "")
ts.plot(lik.comp.CIS[10000:20000], xlab="CIS", main = expression(u),
ylab = "")

ts.plot(lik.comp.RSIS[10000:20000], xlab="RSIS", main = expression(u),
ylab = "")
```

حال با ۵۰۰ تکرار، زنجیر مارکوف مونت‌کارلو را شبیه‌سازی نموده و مقادیر $\hat{Q}(\theta|\tilde{\theta};y)$، $\hat{\sigma}_Q^2$، $t_*\hat{\sigma}_Q/\sqrt{n}$، ACT و $\widehat{ESEJD}$ را به‌دست می‌آوریم.

```
## column 1
qhat.RW <- mean(lik.comp.RW)
qhat.MHIS <- mean(lik.comp.MHIS)
qhat.CIS <- mean(lik.comp.CIS)
qhat.RSIS<- mean(lik.comp.RSIS)

## column 2

xx.RW = lik.comp.RW
nn.RW = length(xx.RW)
```

```
sig.batch.RW = (mcse(xx.RW, size=100)$se)


xx.MHIS = lik.comp.MHIS
nn.MHIS = length(xx.MHIS)
sig.batch.MHIS = (mcse(xx.MHIS, size=100)$se)


xx.CIS = lik.comp.CIS
nn.CIS = length(xx.CIS)
sig.batch.CIS = (mcse(xx.CIS, size=100)$se)



xx.RSIS = lik.comp.RSIS
nn.RSIS = length(xx.RSIS)
sig.batch.RSIS = (mcse(xx.RSIS, size=100)$se)


## column 3

t.star = 1.96

inter.RW  <- t.star*(sig.batch.RW /sqrt(Nsim))
inter.MHIS <- t.star*(sig.batch.MHIS /sqrt(Nsim))
inter.CIS <- t.star*(sig.batch.CIS /sqrt(Nsim))
inter.RSIS <- t.star*(sig.batch.RSIS /sqrt(Nsim))


## column 4

xx.RW = lik.comp.RW
nn.RW = length(xx.RW)
batch.RW = mcse(xx.RW, size=100)$se
```

```
ACT.RW = batch.RW/sqrt(var(xx.RW)/nn.RW)


xx.MHIS = lik.comp.MHIS
nn.MHIS = length(xx.MHIS)
batch.MHIS = mcse(xx.MHIS, size=100)$se
ACT.MHIS = batch.MHIS/sqrt(var(xx.MHIS)/nn.MHIS)


xx.CIS = lik.comp.CIS
nn.CIS = length(xx.CIS)
batch.CIS = mcse(xx.CIS, size=100)$se
ACT.CIS = batch.CIS/sqrt(var(xx.CIS)/nn.CIS)


xx.RSIS = lik.comp.RSIS
nn.RSIS = length(xx.RSIS)
batch.RSIS = mcse(xx.RSIS, size=100)$se
ACT.RSIS = batch.RSIS/sqrt(var(xx.RSIS)/nn.RSIS)

## column 5

#function 2
like.comp <- function(y,X,u.hat,beta=4, sig2=1.5, k=10){
lik.comp <- rep(0,Nsim)
for (i in 1:Nsim){
lik.comp[i] <- L.L.C(y,X,u.hat[i,],beta=4, sig2=1.5, k=10)
}
lik.comp
}

iter = 500

MSE.Q<- function(Nsim = 10000, iter ){
```

```
Q.RW = Q.MHIS = Q.CIS  = Q.RSIS = rep(0,iter)
JD.RW <- JD.MHIS <- JD.CIS <- JD.RSIS <- rep(0,iter)
i=0
for (i in 1:iter){

     result.RW <- RWM.MCMC(Nsim=Nsim)
result.MHIS <- MHIS.MCMC(Nsim=Nsim)
result.CIS <- CIS.MCMC(Nsim=Nsim)
result.RSIS <-  RSIS.MCMC(Nsim=Nsim)

lik.comp.RW <- like.comp(y,X,result.RW$u,beta=4, sig2=1.5, k=10)
lik.comp.MHIS <- like.comp(y,X,result.MHIS$u,beta=4, sig2=1.5, k=10)
lik.comp.CIS <- like.comp(y,X,result.CIS$u,beta=4, sig2=1.5, k=10)
lik.comp.RSIS <- like.comp(y,X,result.RSIS$u,beta=4, sig2=1.5, k=10)
## estimating MSEJD
JD.RW[i] = mean(diff(lik.comp.RW)^2)
JD.MHIS[i] = mean(diff(lik.comp.MHIS)^2)
JD.CIS[i] = mean(diff(lik.comp.CIS)^2)
JD.RSIS[i] = mean(diff(lik.comp.RSIS)^2)
cat("iteration = ", i <- i + 1, "\n")
}
## Estimating ESEJD

      ESEJD.RW = mean(JD.RW)
ESEJD.MHIS = mean(JD.MHIS)
ESEJD.CIS = mean(JD.CIS)
ESEJD.RSIS = mean(JD.RSIS)

## Estimating standard errors of ESEJD estimates

      ESEJD.se.RW = sd(JD.RW)/sqrt(iter)
ESEJD.se.MHIS = sd(JD.MHIS)/sqrt(iter)
ESEJD.se.CIS = sd(JD.CIS)/sqrt(iter)
ESEJD.se.RSIS = sd(JD.RSIS)/sqrt(iter)
```

```
list(
ESEJD.RW = ESEJD.RW, ESEJD.MHIS = ESEJD.MHIS, ESEJD.CIS =ESEJD.CIS,
      ESEJD.RSIS = ESEJD.RSIS,
    ESEJD.se.RW =ESEJD.se.RW, ESEJD.se.MHIS = ESEJD.se.MHIS ,ESEJD.se.CIS = ESEJD.s
      ESEJD.se.RSIS =ESEJD.se.RSIS)
}

output = MSE.Q(Nsim=10000 , iter=2)
output
```

مراجع

[۱] پاشا ع، (۱۳۸۰) " فرآیندهای تصادفی (رشته آمار)" ، انتشارات دانشگاه پیام‌نور، تهران.

[2] Atchadé, Y. F. (2011) Kernel Estimators of Asymptotic Variance for Adaptive Markov Chain Monte Carlo, *The Annals of Statistics*, **39**, 990-1011.

[3] Bédard, M. and Rosenthal, J. S. (2008) Optimal Scaling of Metropolis Algorithms: Heading Toward General Target Distributions, *Canadian Journal of Statistics*, **36**, 483-503.

[4] Birkhoff, G. D. (1942) What is the Ergodic Theorem?, *American Mathematical Monthly*, **49**, 222-226.

[5] Booth, J. G. and Hobert, J. P. (1999) Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**, 265–285.

[6] Breslow, N. E. and Clayton, D. G. (1993) Approximate Inference in Generalized Linear Mixed Model,*Journal of the American Statistical Association*, **88**, 9-25

[7] Caffo, B. S., Jank, W. and Jones, G. L. (2005) Ascentbased Monte Carlo Expectation-Maximization, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 235–251.

[8] Casella, G. and Berger, R. L. (2002) Statistical inference , **2**. Pacific Grove, CA: Duxbury.

[9] Casella, G. and George, I. E. (1992) Explaining the Gibbs Sampler, *The American Statistician*, **46**, 167-174.

[10] Christensen, O. F., Møller, J. and Waagepetersen, R. P. (2001) Geometric Ergodicity of Metropolis-Hastings Algorithms for Conditional Simulation in Generalized Linear Mixed Models, *Methodology and Computing in Applied Probability*, **3**, 309-327.

[11] Coull, B. A., Hobert, J. P., Ryan, L. M. and Holmes, L. B. (2001) Crossed Random Effect Models for Multiple Outcomes in a Study of Teratogenesis, *Journal of the American Statistical Association*, **96**, 1194–1204.

[12] Diaconis, P., Khare, K., and Saloff-Coste, L. (2008) Gibbs Sampling, Exponential Families and Orthogonal Polynomials, *Statistical Science*, **23**, 151-178.

[13] Doss, H. and Hobert, J. P. (2012) Estimation of Bayes Factors in a Class of Hierarchical Random Effects Models Using a Geometrically Ergodic MCMC Algorithm, *Journal of Computational and Graphical Statistics*.

[14] Doss, H. and Hobert, J. P. (2010) Estimation of Bayes Factors in a Class of Hierarchical Random Effects Models Using a Geometrically Ergodic MCMC Algorithm, *Journal of Computational and Graphical Statistics*, **19**, 295-312.

[15] Flegal, J. M., Haran, M. and Jones, G. L. (2008) Markov Chain Monte Carlo: Can We Trust the Third Significant Figure?, *Statistical Science*, **23**, 250–260.

[16] Flegal, J. M. and Jones, G. L. (2010) Batch Means and Spectral Variance Estimators in Markov Chain Monte Carlo *The Annals of Statistics*, **38**, 1034–1070.

[17] Flegal, J. M. and Jones, G. L. (2011) Implementing Markov Chain Monte Carlo: Estimating with Confidence, *In Handbook of Markov Chain Monte Carlo*, (S. P. Brooks, A. Gelman, G. L. Jones and X. L. Meng, eds.). CRC Press, Boca Raton, FL.

[18] Fort, G., Moulines, E., Roberts, G. O. and Rosenthal, J. S. (2003) On the Geometric Ergodicity of Hybrid Samplers, *Journal of Applied Probability*, 123-146.

[19] Gelfand, A. and Smith, A. (1990) Sampling Based Approaches to Calculating Marginal Densities, *Journals American Statistical Association*, **85**, 398-409.

[20] Gelman, A., Kass, R.E., Carlin, B.P. and Neal, R. (1998) Markov Chain Monte Carlo in Practice: a Roundtable Discussion, *The American Statistician*, **52**, 93-100

[21] Gelman, A. and Rubin, D.B. (1992) Inference From Iterarative Simulation Using Multiple Sequence, *Statistical Science*, **7**, 457-511

[22] Geman, S. and Geman, D. (1984) Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

[23] Geweke, J. (1992) Evaluating the Accuacy of Sampling-Based Approaches to Calculating Posterior Momments, *In Bayesian Statistics*, (Eds JM Bernado, JO Berger, AP Dawid and AFM Smith), **4**, 169-193.

[24] Geyer, C. J. (1999) Likelihood Inference for Spatial Point Processes, *Stochastic Geometry: Likelihood and Computation*, **80**, 79-140.

[25] Hastings, W. (1970) Monte Carlo Sampling Methods Using Markov Chains and their Application, *Biometrika*, **57**, 97-109.

[26] Heidelberger, P. and Welch, P. D. (1983) Simulation Run Length Control in the Presence of an Initial Transient, *Operations Research*, **31**, 1109-1144.

[27] Hesterberg, T. C. (2003) Advances in importance sampling *(Doctoral dissertation, Stanford University)*.

[28] Hobert, J. P. (2000) Hierarchical Models: A Current Computational Perspective, *Journal of the American Statistical Association*, **95**, 1312–1316.

[29] Hobert, J. P. (2011) The Data Augmentation Algorithm: Theory and Methodology, *In Handbook of Markov Chain Monte Carlo* (S. P. Brooks, A. Gelman, G. L. Jones and X. L. Meng, eds.). CRC Press, Boca Raton, FL.

[30] Hobert, J. P. and Geyer, C. J. (1998) Geometric Ergodicity of Gibbs and Block Gibbs Samplers for a Hierarchical Random Effects Model, *Journal of Multivariate Analysis*, **67**, 414-430.

[31] Jarner, S. F. and Hansen, E. (2000) Geometric Ergodicity of Metropolis Algorithms, *Stochastic Processes and Their Applications*, **85**, 341-361.

[32] Johnson, A. A. (2009) Geometric ergodicity of Gibbs Samplers *(Doctoral Dissertation, University of Minnesota)*.

[33] Johnson, A. A. and Jones, G. L. (2010) Gibbs Sampling for a Bayesian Hierarchical General Linear Model, *Electronic Journal of Statistics*, **4**, 313-333.

[34] Johnson, A. A. and Jones, G. L. (2015) Geometric Ergodicity of Random Scan Gibbs Samplers for Hierarchical One-Way Random Effects Models, *Journal of Multivariate Analysis*, **140**, 325-342.

[35] Johnson, L. T. and Geyer, C. J. (2012) Variable Transformation to Obtain Geometric Ergodicity in the Random-Walk Metropolis Algorithm, *The Annals of Statistics*, **40**, 3050-3076.

[36] Jones, G. L. (2004) On the Markov chain central limit theorem, *Probability Surveys*.

[37] Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006) Fixed-Width Output Analysis for Markov Chain Monte Carlo, *Journal of the American Statistical Association*, **101**, 1537–1547.

[38] Jones, G. L. and Hobert, J. P. (2001) Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo,*Statist. Sci*, **16**, 312–334.

[39] Jones, G. L., Roberts, G. O. and Rosenthal, J. S. (2014) Convergence of Conditional Metropolis–Hastings Samplers, *Advances in Applied Probability*, **46**, 422-445.

[40] Latuszynski, K., Roberts, G. O. and Rosenthal, J. S. (2013) Adaptive Gibbs Samplers and Related MCMC Methods,*The Annals of Applied Probability*,**23**, 66–98.

[41] Lee, K. J., Jones, G. L., Caffo, B. S. and Bassett, S. (2013) Spatial Bayesian Variable Selection Models on Functional Magnetic Resonance Imaging Time-Series Data. Preprint.

[42] Mcculloch, C. E. (1997) Maximum Likelihood Algorithms for Generalized Linear Mixed Models, *Journal of the American statistical Association*, **92**, 162–170.

[43] Mcculloch, C. E., Nelder, Searle, S. R. (2001) Generalized Linear and Mixed Models, *John Wiley and Sons, New York*.

[44] Mccullagh P. and Nelder, J. A. (1989) Generalized Linear Models,*London*.

[45] Mengersen, K. L. and Tweedie, R. L. (1996) Rates of Convergence of the Hastings and Metropolis Algorithms, *The Annals of Statistics*, **24**, 101-121.

[46] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, **21**, 1087-1092.

[47] Meyn, S. P. and Tweedie, R. L. (1993) *Markov Chains and Stochastic Stability*, Springer, London.

[48] Meyn, S. P. and Tweedie, R. L. (1993) Markov Chains and Stochastic Stability, Communication and Control Engineering Series, *Springer-Verlag London Ltd., London*, **1**, 993.

[49] Meyn, S. P. and Tweedie, R. L. (1994) Computable Bounds for Geometric Convergence rates of Markov Chains, *The Annals of Applied Probability*, **4**, 981–1011.

[50] Neal, P. and Roberts, G. (2006) Optimal Scaling for Partially Updating MCMC Algorithms, *The Annals of Applied Probability*, **16**, 475-515.

[51] Neath, R. C. (2013) On Convergence Properties of the Monte Carlo EM Algorithm, *In Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton* (G. L. Jones and X. Shen, eds.). To appear.

[52] Nelder, J. A. and Wedderburn, R.W.M. (1972) Generalized Linear Models, *Journal of the Royal Statistical Society*, **135**, 370-384.

[53] Nummelin, E. (1984) General Irreducible Markov Chains and non-negative Operators, *Cambridge University Press. MR776608*.

[54] Pinherio, J. C. and Bates, D. M. (1995) Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Models, Journal of Computational and Graphical Statistics, **4**, 12-35.

[55] Robert, C. and Casella, G. (2004) Monte Carlo Statistical Methods, 2nd edn. *Springer, New York*.

[56] Robert C. and Casella, G. (2010) Introducing Monte Carlo Methods with R, *Springer, New York*.

[57] Robert, C. and Casella, G. (2011) A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data, *Statistical Science*, 102-115.

[58] Robert, J. (2011) The Data Augmentation Algorithm: Theory and Methodology. In Handbook of Markov Chain Monte Carlo (S. P. Brooks, ANGelman, G. L. Jones and X. L. Meng, eds.), *CRC Press, Boca Raton, FL*

[59] Robert, C. P. (1995) Convergence Control Methods for Markov Chain Monte Carlo Algorithms *Statistical Science*, 231-253.

[60] Robert, C. (2007) The Bayesian Choice: from Decision-Theoretic Foundations to Computational Implementation. *Springer Science and Business Media*.

[61] Roberts, G. O. and Rosenthal, J. S. (1997) Geometric Ergodicity and Hybrid Markov Chains, *Electron. Comm. Probab*, **2**, 13-25.

[62] Roberts, G. O. and Rosenthal, J. S. (2001) Markov Chains and De-initializing Processes, *Scandinavian Journal of Statistics*, **28**, 489-504.

[63] Roberts, G. O. and Rosenthal, J. S. (2004) General State Space Markov Chains and MCMC Algorithms,*Probability Surveys*, **1** , 20–71.

[64] Roberts, G. O. and Sahu, S. K. (1997) Updating Schemes, Correlation Structure, Blocking and Parameterization for the Gibbs Sampler, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59**, 291-317.

[65] Roberts, G. O. and Tweedie, R. L. (1996) Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms, *Biometrika*, **83**, 95-110.

[66] Rom´an, J. C. (2012) Convergence Analysis of Block Gibbs Samplers for Bayesian General Linear Mixed Models, *Ph.D. Thesis, Dept. Statistics, Univ. Florida*

[67] Rom´an, J. C. and Hobert, J. P. (2012) Convergence Analysis of the Gibbs Sampler for Bayesian General Linear Mixed Models with Improper Priors, *The Annals of Statistics*, **40**, 2823–2849.

[68] Rosenthal, J. S. (2011) Optimal Proposal Distributions and Adaptive MCMC, *In Handbook of Markov Chain Monte Carlo (S. P. Brooks, A. Gelman, G. L. Jones and X. L. Meng, eds.). CRC Press, Boca Raton, FL.*

[69] Roy, V. and Hobert, J. P. (2007) Convergence Rates and Asymptotic Standard Errors for Markov Chain Monte Carlo Algorithms for Bayesian Probit Regression, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**, 607–623.

[70] Tan, A., Jones, G. L. and Hobert, J. P. (2013) On the Geometric Ergodicity of Two-Variable Gibbs Samplers, *In Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton (G. L. Jones and X. Shen, eds.). To appear.*

[71] Tanner, M. A. and Wong, W. H. (1987) The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, **82**, 528-540.

[72] Tierney, L. (1994) Markov Chains for Exploring Posterior Distributions (with Discussion), *the Annals of Statistics*, **22**, 1701–1762.

[73] Tierney, L. and Kadane, J. B. (1986) Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the american statistical association*, **81**, 82-86.

واژه‌نامه فارسی به انگلیسی

اثرات تصادفی ..... Random effects

اثرات ثابت ..... Fixed effects

ارگودیک هریس ..... Harris ergodic

ارگودیک هندسی ..... Geometric ergodic

ارگودیک یکنواخت ..... Uniform ergodic

اسکن تصادفی ..... Random scan

الگوریتم متروپولیس–هستینگز تعدیل‌شده لانگوین ..... Metropolis-Hastings Adjusted Langevin algorithm

الگوریتم متروپولیس–هستینگز درون گیبس ..... Metropolis-Within-Gibbs algorithm

الگوریتم متروپولیس–هستینگز قدم‌زدن تصادفی ..... Random Walk Metropolis-Hastigs Algorithm algorithm

الگوریتم متروپولیس–هستینگز مستقل ..... Independent Metropolis-Hastings algorithm

الگوریتم‌های مونت‌کارلو زنجیر مارکوف سازوار ..... Adaptive MCMC algorithms

آمیختگی دنباله‌ای ..... Sequence mixing

آمیختن ..... Mixing

اندازه ..... Measure

بازگشت‌پذیری ..... Reversiblity

برگشت‌پذیر هریس ..... Harris recurrent

به‌روز رسانی مولفه به مولفه ..... Component-wise updating

پروبیت ..... Probit

پیشگوی خطی ..... Linear preditor

تابع پیوند ..... Link function

تابع درستنمایی داده کامل ..... Complete-data log-likelihood

تحویل‌ناپذیری ..... Irreduciblity

ترکیب ..... Composition

ترکیب ..... Combine
تقریب لاپلاس ..... Laplace approximation
توزیع پیشنهادی ..... Proposal distribution
توزیع پیشین ..... Perior destribution
توزیع مانا ..... Invariant distribution
توزیع‌های خانواده نمایی ..... Exponential family distributions
ثابت نرمال‌ساز ..... Normalizing constant
چگالی انتقال مارکوف ..... Mrkov transition density
چگالی پسین ..... Posterior density
حجم نمونه موثر ..... Effective Sample Size
داده‌های طولی ..... Longitudinal data
دلتای دیراک ..... Dirac's delta
دنباله تصادفی ..... Random sequence
روش‌های طیفی ..... Spectral methods
روش‌های مربع‌بندی ..... Quadrature methods
مونت‌کارلو زنجیر مارکوف ..... Markov Chain Monte Carlo
شرط تعادل دقیق ..... Detailed balance condition
شرطی کامل ..... Full conditional
فاصله تغییرات کل ..... Total variation distance
فرآیند تصادفی ..... Stochastic process
قضیه ارگودیک ..... Ergodic theorem
قضیه حد مرکزی ..... Central Limited Theorem
کاوش در همسایگی ..... Local exploration
لوجیت ..... Logit
متروپولیس-هستینگز ..... Metropolis Hasting
متغیرهای پنهان ..... Latent variables
مدل آمیخته خطی ..... Linear mixed models
مدل خطی ..... Linear Model
مدل خطی تعمیم‌یافته ..... Generalized Linear Model
مربع امید ریاضی فاصله اقلیدسی پرش ..... Expected Square Euclidean Jump Distance
مونت‌کارلوی امید ریاضی ماکسیمم‌سازی ..... MC expection maximination
میانگین دسته‌ای ..... Batch means

میانگین مربع خطا ..... Mean Square Error
میانگین مربع فاصله اقلیدسی پرش ..... Mean Square Euclidean Jump Distance
نادوره‌ای ..... Aperiodic
نمونه‌گیر رد و پذیرش ..... Reject-accept sampling
نمونه‌گیر گیبز ..... Gibbs Sampler
نمونه‌گیری مهم ..... Importance sampling
هسته انتقال ..... Transition kernel
هستینگز ..... Hastings

واژه‌نامه انگلیسی به فارسی

الگوریتم‌های مونت‌کارلو زنجیر مارکوف سازوار ..... Adaptive MCMC algorithms
نادوره‌ای ..... Aperiodic
همبستگی یکپارچه زمان ..... Autocorrelation Time
میانگین دسته‌ای ..... Batch means
قضیه حد مرکزی ..... Central Limited Theorem
ترکیب ..... Combine
درستنمایی داده کامل ..... Complete-data log-likelihood
به‌روز رسانی مولفه به مولفه ..... Component-wise updating
ترکیب ..... Composition
شرط تعادل دقیق ..... Detailed balance condition
دلتای دیراک ..... Dirac's delta
حجم نمونه موثر ..... Effective Sample Size
قضیه ارگودیک ..... Ergodic theorem
مربع امید ریاضی فاصله اقلیدسی پرش ..... Expected Square Euclidean Jump Distance
توزیع‌های خانواده نمایی ..... Exponential family distributions
اثرات ثابت ..... Fixed effects
شرطی کامل ..... Full conditional
مدل خطی تعمیم‌یافته ..... Generalized Linear Models
ارگودیک هندسی ..... Geometric ergodic
نمونه‌گیر گیبز ..... Gibbs Sampler
ارگودیک هریس ..... Harris ergodic
برگشت‌پذیر هریس ..... Harris recurrent
هستینگز ..... Hastings
نمونه‌گیری مهم ..... Importance sampling
الگوریتم متروپولیس-هستینگز مستقل ..... Independent Metropolis-Hastings algorithm

توزیع مانا ..... Invariant distribution
تحویل‌ناپذیری ..... Irreduciblity
تقریب لاپلاس ..... Laplace approximation
متغیرهای پنهان ..... Latent variables
مدل آمیخته خطی ..... Linear mixed models
مدل خطی ..... Linear Model
پیشگوی خطی ..... Linear preditor
تابع پیوند ..... Link function
کاوش در همسایگی ..... Local exploration
لوجیت ..... Logit
داده‌های طولی ..... Longitudinal data
مونت‌کارلو زنجیر مارکوف ..... Markov Chain Monte Carlo
چگالی انتقال مارکوف ..... Markov transition density
میانگین مربع خطا ..... Mean Square Error
میانگین مربع فاصله اقلیدسی پرش ..... Mean Square Euclidean Jump Distance
اندازه ..... Measure
متروپولیس-هستینگز ..... Metropolis Hasting
الگوریتم متروپولیس-هستینگز تعدیل‌شده لانگوین ..... Metropolis-Hastings Adjusted Langevin algorithm
الگوریتم متروپولیس-هستینگز درون گیبس ..... Metropolis-Within-Gibbs algorithm
آمیختن ..... Mixing
مونت‌کارلوی امید ریاضی ماکسیمم‌سازی ..... MC expection maximination
ثابت نرمال‌ساز ..... Normalizing constant
توزیع پیشین ..... Perior destribution
چگالی پسین ..... Posterior density
پروبیت ..... Probit
توزیع پیشنهادی ..... Proposal distribution
روش‌های مربع‌بندی ..... Quadrature methods
اثرات تصادفی ..... Random effects
اسکن تصادفی ..... Random scan
دنباله تصادفی ..... Random sequence
الگوریتم متروپولیس-هستینگز قدم‌زدن تصادفی ..... Random Walk Metropolis-Hastigs algorithm

Algorithm

نمونه‌گیر رد و پذیرش ..... Reject-accept sampling

بازگشت‌پذیری ..... Reversiblity

آمیختگی دنباله‌ای ..... Sequence mixing

روش‌های طیفی ..... Spectral methods

فرآیند تصادفی ..... Stochastic process

فاصله تغییرات کل ..... Total variation distance

هسته انتقال ..... Transition kernel

ارگودیک یکنواخت ..... Uniform ergodic

نمایه

Abstract

Markov Chain Monte Carlo (MCMC) methods are among the popular sampling methods for approximating posterior distribution of complex models with high dimension. In situations where the dimension of posterior distribution is high or/and its components are weakly correlated, it is common practice to update the simulation one variable (or subblock of variables) at a time, rather than conduct a single full-dimensional update; this is what is called component-wise updating. for example, Gibbs sampling and Metropolis-Hastings-within-Gibbs algorithms are component-wise algorithms. There are different strategies for combining component-wise updates. While these strategies can ease MCMC implementation and produce superior empirical performance compared to full-dimensional updates, the theoretical convergence properties of the associated Markov chains have received limited attention. We review conditions under which some component-wise Markov chains converge to the stationary distribution at a geometric rate. The particular attention is on the connections between the convergence rate of the various component-wise strategies. We assess the theoretical results of the thesis by two simulation examples.

Keywords: Geometric ergodicity, uniform ergodicity, Markov chain, Monte Carlo, Gibbs sampler, Metropolis-within-Gibbs, random scan, convergence rate.

Shahrood University of Technology

Faculty Of Mathematical Sciences

Component-Wise Markov Chain Monte Carlo: Uniform and Geometric Ergodicity under Mixing and Composition

By: Fatemeh Khodabakhshi Palandi

Supervisor

Dr. Negar Eghbal

Advisor

Dr. Hossein Baghishani

Septembear 2016