



دانشگاه شاهرود

دانشکده علوم ریاضی
گروه ریاضی کاربردی

پایان نامه

برای دریافت درجه کارشناسی ارشد در رشته
آمار، گرایش آمار ریاضی

عنوان

محاسبات بیزی تقریبی: روش‌ها و کاربردها

استاد راهنما

دکتر حسین باغیشنی

استاد مشاور

دکتر نگار اقبال

دانشجو

سعید انجمنی

بهمن ۱۳۹۳

تقدیم بہ پدر و مادر عزیزم

نخستین سپاس و ستایش از آن خداوندی است که بنده کوچکش را در دریای بیکران
اندیشه، قطره‌ای ساخت تا وسعت آن را از دریچه اندیشه‌های ناب آموزگارانی
بزرگ به تماشا نشیند.

اکنون که در سایه‌سار بنده‌نوازی‌هایش، پایان‌نامه حاضر به انجام رسیده است، بر خود لازم می‌دانم تا
مراتب سپاس را از بزرگوارانی به جا آورم که اگر دست یاری‌گرشان نبود، هرگز این پایان‌نامه به انجام
نمی‌رسید.

ابتدا از استاد گران‌قدم آقای دکتر باغی‌شینی که زحمت راهنمایی این پایان‌نامه را بر عهده داشتند،
کمال سپاس را دارم. از استاد عالی‌قدم سرکار خانم دکتر اقبال که زحمت مشاوره این پایان‌نامه را
متحمل شدند، صمیمانه تشکر می‌کنم. بر خود لازم می‌دانم که از اساتید محترم داور، جناب آقایان دکتر
شاهسونی و دکتر ربیعی، و همچنین از تمام اساتید فرهیخته گروه آمار دانشگاه شاهرود، تشکر و قدردانی
نمایم.

سپاس آخر را به مهربان‌ترین همراهان زندگیم، به پدر و مادر عزیزم تقدیم می‌کنم که حضورشان در
فضای زندگیم مصداق بی‌ریای سخاوت بوده است.

سعدا محمدی
بهمن ۱۳۹۳

تعمدنامه

اینجانب سعید افچنگی دانشجوی کارشناسی ارشد رشته آمار دانشکده علوم ریاضی دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان محاسبات بیزی تقریبی: روش‌ها و کاربردها، تحت راهنمایی دکتر حسین باغیشنی 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، تاکنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه شاهرود متعلق دارد، و مقالات مستخرج با نام “ دانشگاه شاهرود “ یا “ Shahrood University “ 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.

سعید افچنگی
بهمن ۱۳۹۳

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 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

روش‌های محاسبات بیزی تقریبی، به‌عنوان تکنیکی از روش‌های نمونه‌گیری مبتنی بر درست‌نمایی-آزاد شناخته می‌شوند. این روش‌ها، در ده سال اخیر، به‌عنوان راهکاری رضایت‌بخش برای حل مشکل استنباط بیزی مدل‌های با توابع درست‌نمایی پیچیده و رام‌نشدنی، معرفی شده‌اند. محاسبات بیزی تقریبی، برای اولین بار در ژنتیک و سپس در طیف وسیع‌تری از موقعیت‌های کاربردی مورد استفاده قرار گرفته شد. در طی چند سال گذشته، هر دو جنبه کاربرد و مباحث نظری این روش‌ها مورد نظر محققین در آمار و سایر رشته‌ها قرار گرفته است. با توجه به استقبال گسترده و بی‌سابقه از این روش‌ها، در این پایان‌نامه به معرفی الگوریتم‌های مختلف آن‌ها، بررسی برخی از مباحث نظری و نمایش چند کاربرد آن‌ها در مثال‌های واقعی و شبیه‌سازی، می‌پردازیم.

کلمات کلیدی

روش‌های درست‌نمایی-آزاد، آمار بیزی، روش ABC، ABC نوفه‌دار.

پیشگفتار

در بسیاری از کاربردهای آماری شامل استنباط در مدل‌های تصادفی پیچیده، مانند سیستم‌های تصادفی پویا، مشکل اصلی محاسبه تابع درست‌نمایی است. در چنین کاربردهایی، محاسبه تابع درست‌نمایی پیچیده و حتی ناممکن است. در کنار این پیچیدگی، با شرط معلوم بودن پارامترهای مدل، شبیه‌سازی از تابع درست‌نمایی مدل کار ساده‌ای است. روش محاسبات بیزی تقریبی (ABC) در چنین شرایطی به‌کار می‌رود (بیومونت و همکاران، ۲۰۰۲). این روش به‌جای محاسبه تابع درست‌نمایی از مرحله‌ای که شامل شبیه‌سازی داده‌ها به ازای پارامترهای مختلف است، استفاده می‌کند. سپس به مقایسه داده‌ها یا آماره‌های خلاصه‌شده‌ای از آن‌ها برای داده‌های مشاهده‌شده و شبیه‌سازی شده می‌پردازد، به طوری که اگر به هم نزدیک باشند، مقدار پارامتری که داده‌ها تحت آن شبیه‌سازی شده‌اند را به عنوان مقداری از توزیع پسین، می‌پذیرد. این روش در طی سال‌های اخیر، از استقبال گسترده‌ای، به‌ویژه در حوزه‌های کاربردی، برخوردار شده است. به همین دلیل در این پایان‌نامه بر آن شدیم تا الگوریتم‌های مختلف این روش، برخی از مباحث نظری پیرامون آن و چند کاربرد از آن را معرفی کنیم و مورد بحث قرار دهیم. مطالب پایان‌نامه، به اختصار، شامل موارد زیر می‌باشد:

- در فصل اول، مفاهیم و تعاریف اولیه را معرفی می‌کنیم که شامل الگوریتم‌های نمونه‌گیری مونت کارلو می‌باشد. سپس اجزای لازم برای روش‌های محاسبات بیزی تقریبی را به همراه نگاهی بر روش‌های درست‌نمایی-آزاد بیان می‌کنیم.
- در فصل دوم، ابتدا تاریخچه‌ای از روش ABC را ارائه می‌دهیم. در ادامه، سه نسخه اصلی روش‌های ABC را به‌طور مفصل مورد بررسی قرار می‌دهیم و الگوریتم‌های آن‌ها را بیان می‌کنیم. سپس با ارائه یک مثال شبیه‌سازی، عملکرد آن را ارزیابی می‌کنیم.
- در فصل سوم، برخی از مطالب نظری پیرامون روش ABC را مورد مطالعه قرار می‌دهیم. در این فصل، به خطاهای ایجادشده در برآورد پسین اشاره می‌کنیم، و سعی در کاهش این خطاها خواهیم داشت. این تلاش منجر به معرفی مقادیر بهینه برای مولفه‌های روش ABC، از جمله استفاده از میانگین توزیع پسین برای آماره خلاصه‌شده، می‌گردد. در ادامه، ویژگی‌های مجانبی برآوردگرهای بیزی حاصل از ABC، شامل نرخ همگرایی، را در یک قضیه کلی ۶.۳.۳ بیان و اثبات می‌کنیم.
- در فصل چهارم، کاربرد و نحوه استنباط و شبیه‌سازی با استفاده از روش ABC و نرم‌افزار R را در دو مثال کاربردی نمایش می‌دهیم.
- در این پایان‌نامه، پیوست شامل برخی تعاریف لازم و پیوست B شامل برخی از کدهای نوشته‌شده در محیط نرم‌افزار R، برای بازتولید مثال‌های موجود در پایان‌نامه، است.

فهرست نشانه‌ها و نمادها

Y	متغیر تصادفی
\mathbf{Y}	بردار تصادفی
$\pi(\theta y)$	توزیع پسین
$\pi(\theta)$	توزیع پیشین
$\pi(y \theta)$	تابع درست‌نمایی
$C(\theta)$	ثابت نرمال‌ساز
Θ	فضای پارامتر
RS	نمونه‌گیری رد
$\pi_*(\theta)$	توزیع پیشنهادی
IS	نمونه‌گیری نقاط مهم
\hat{J}_π	برآوردگر مونت کارلوی کلاسیک
$q(\theta)$	تابع ابزار یا پوش
\hat{J}_{IS}	برآوردگر مونت کارلوی IS
w_i	وزن‌های نمونه‌گیری IS
$MCMC$	نمونه‌گیری مونت کارلو زنجیر مارکوفی
$k(\cdot \cdot)$	هسته انتقال مارکوف
$\ \cdot\ $	تابع نرم
$\mathcal{K}(\cdot)$	تابع هسته
$\Phi(y)$	چندک نرمال استاندارد
$mad(y)$	انحراف مطلق میانه
med	میانه
T	تعداد تکرارها (نمونه‌ها)
$\rho(\cdot)$	تابع فاصله
ϵ	سطح تحمل
$S(\cdot)$	آماره خلاصه‌شده
$\mathcal{I}(\cdot)$	تابع نشانگر
ABC	محاسبات بیزی تقریبی
$RS - ABC$	نمونه‌گیری رد ABC

$IS - ABC$	نمونه‌گیری نقاط مهم ABC
$MCMC - ABC$	نمونه‌گیری مونت کارلو زنجیر مارکوفی ABC
T_{acc}	تعداد تکرارهای پذیرفته شده
T_{ess}	حجم نمونه موثر
$L(\theta_0, \hat{\theta}, \cdot)$	تابع زیان توان دوم خطا
MSE	میانگین توان دوم خطا
$Cost$	هزینه
$c(\cdot)$	تابع هزینه
$bias$	اریبی
∂	عملگر مشتق
$tr()$	اثر ماتریس
Σ	ماتریس کواریانس

فهرست مطالب

ر	لیست تصاویر
ز	لیست جداول
۱	۱ تعاریف و مفاهیم اولیه
۱	۱.۱ مقدمه
۲	۲.۱ استنباط آماری
۲	۳.۱ دیدگاه بسامدی
۴	۴.۱ دیدگاه بیزی
۵	۱.۴.۱ توزیع پسین
۶	۵.۱ روش‌های نمونه‌گیری
۶	۱.۵.۱ نمونه‌گیری رد
۸	۲.۵.۱ نمونه‌گیری نقاط مهم
۱۰	۳.۵.۱ نمونه‌گیری مونت کارلوی زنجیر مارکوفی
۱۲	۶.۱ نرم‌های آماری
۱۳	۱.۶.۱ انواع نرم‌ها
۱۳	۷.۱ هسته‌ها
۱۴	۸.۱ مدل‌های تصادفی پیچیده
۱۶	۱.۸.۱ مدل $k-g$
۱۷	۲.۸.۱ مدل‌های α -پایدار
۱۷	۳.۸.۱ چگالی توام میانه و انحراف مطلق میانه
۱۷	۹.۱ روش‌های درست‌نمایی آزاد
۱۹	۱۰.۱ تولید تحقق‌هایی از مدل
۱۹	۱.۱۰.۱ آماره‌های خلاصه شده
۲۰	۲.۱۰.۱ تابع فاصله (نرم)
۲۰	۳.۱۰.۱ تصمیم و تابع وزنی
۲۱	۴.۱۰.۱ سطح تحمل

۲۳	۲ محاسبات بیزی تقریبی
۲۳	۱.۲ مقدمه
۲۴	۱.۱.۲ انگیزه پیدایش روش‌های ABC
۲۴	۲.۲ ABC چیست؟
۲۵	۱.۲.۲ سطوح رامنش‌دنی در تابع درست‌نمایی
۲۶	۲.۲.۲ کارکرد روش ABC
۲۸	۳.۲ الگوریتم‌های مختلف ABC
۲۹	۱.۳.۲ نمونه‌گیری رد-ABC
۳۵	۲.۳.۲ نمونه‌گیری نقاط مهم ABC
۳۶	۳.۳.۲ نمونه‌گیری مونت کارلوی زنجیر مارکوفی ABC
۴۰	۴.۲ عیوب و راه‌حل‌ها
۴۱	۳ چارچوب نظری روش‌های ABC
۴۲	۱.۳ کالیبدن و ABC نوفه‌دار
۴۲	۱.۱.۳ کالیبدن چیست؟
۴۳	۲.۱.۳ الگوریتم ABC نوفه‌دار
۴۶	۲.۳ سه سطح تقریب و سه خطا برای برآورد پسین پارامترها
۴۶	۱.۲.۳ خطای مونت کارلو ABC
۵۱	۲.۲.۳ خطای تقریب پسین در استفاده از آماره‌ها (غیربسنده)
۵۴	۳.۲.۳ خطای تقریب پسین در استفاده از آستانه تحمل (یا هسته) و نرم
۵۶	۳.۳ نرخ همگرایی
۵۸	۱.۳.۳ نتایج نظری نرخ همگرایی
۶۷	۴ کاربردها
۶۷	۱.۴ مقدمه
۶۷	۱.۱.۴ توزیع وایبل
۶۸	۲.۱.۴ مدل MA(q)
۷۱	۳.۱.۴ مدل‌های زیست‌شناسی پویا
۸۱	آ نمادها و تعریف‌ها
۸۳	ب کد برنامه R برای اجرای فصل دوم و چهارم
۹۶	مراجع
۱۰۲	واژه‌نامه فارسی به انگلیسی

واژه‌نامه انگلیسی به فارسی

۱۰۵

لیست تصاویر

۱۰۲	خروجی برای ۲۰ تکرار الگوریتم ۴.۲ بر اساس مدل نرمال است. نمودار سمت چپ توزیع تجمعی تجربی داده واقعی و شبیه‌سازی شده و سمت راست مقادیر پذیرفته را نشان می‌دهد.	۳۲
۲۰۲	خروجی الگوریتم ۱۰.۲ بر اساس ۱۰۰۰ تکرار از توزیع نرمال است. سطرها از بالا به پایین، به ترتیب متناظر با مقادیر ۰/۰۱، ۰/۰۵، ۰/۱ و ۰/۲۵ برای ϵ هستند. در تمام نمودارها محور افقی θ ها هستند.	۳۳
۳۰۲	نمودارهای مشابه شکل ۲۰.۲ که بر اساس آماره بسنده $S(y) = \bar{y}$ نتیجه شده‌اند	۳۴
۱۰۳	نمودار چگالی پسین حاشیه‌ای پارامترهای (θ_1, θ_2) بر اساس آماره‌های خلاصه‌شده میانه و mad (دو هیستوگرام بالایی) و میانگین و واریانس (دو هیستوگرام پایینی).	۵۴
۱۰۴	پارامترهای پذیرفته شده بر اساس چندک‌های $(1/5, 1, 4/5)$ $\epsilon =$ درصد از فاصله مورد نظر، و آماره خلاصه شده میانگین و انحراف معیار	۶۸
۲۰۴	خروجی الگوریتم پذیرش رد بر اساس تحمل ϵ برابر با ۰/۱ درصد از $N = 10^6$ داده‌های شبیه‌سازی پذیرفته شده	۷۰
۳۰۴	پراکندگی دو نمونه ABC را بر اساس فاصله خام و فاصله اتوکواریانس محاسبه شده است که با استفاده از چندک‌های متفاوت ϵ (آبی برای ۱٪، قرمز برای ۰/۱٪ و زرد برای ۰/۱٪) نشان داده شده است. سطوح چگالی پسین به رنگ مشکی است.	۷۱
۴۰۴	توزیع نمونه‌ای ABC را با استفاده از چندک‌های متفاوت نقطه نقطه برای ۱۰٪، خط تیره برای ۱٪ و خط توپر برای ۰/۱٪ درصد، که در مقایسه با چگالی حاشیه‌ای واقعی نشان داده شده است. داده‌ها مشابه نمودار ۳۰۴ است.	۷۲
۵۰۴	نمودار داده $LVperfect$ نسبت به زمان t است که خط چین بیانگر داده شکارچی و خط پر بیانگر داده شکار است.	۷۶
۶۰۴	نمودار هیستوگرام پارامترهای شبیه‌سازی شده بر اساس θ و $\log(\theta)$	۷۷
۷۰۴	نمودار هیستوگرام پارامترهای شبیه‌سازی شده بر اساس θ و $\log(\theta)$ با استفاده از بسته <i>parallel</i>	۷۸
۸۰۴	نمودار هیستوگرام پارامترهای شبیه‌سازی شده بر اساس θ و $\log(\theta)$	۷۸

لیست جداول

۷	الگوریتم نمونه‌گیری رد	۱.۱
۱۰	الگوریتم نمونه‌گیری نقاط مهم	۲.۱
۱۲	الگوریتم نمونه‌گیری MCMC	۳.۱
۱۵	توابع هسته آماری رایج	۴.۱
۲۷	نمونه‌گیری از توزیع پسین واقعی با الگوریتم ABC	۱.۲
۲۷	نمونه‌گیری از توزیع پسین با الگوریتم ABC	۲.۲
۲۸	نمونه‌گیری از توزیع پسین با الگوریتم ABC و بر حسب آماره خلاصه‌شده	۳.۲
۳۰	نمونه‌گیری RS-ABC	۴.۲
۳۶	نمونه‌گیری IS-ABC	۵.۲
۳۷	نمونه‌گیری MCMC-ABC	۶.۲
۴۴	افزودن مرحله آماره خلاصه‌شده تصادفی در الگوریتم ABC نوفه‌دار	۱.۳
۷۴	با توجه به عبارات (۶.۴)، (۷.۴) و (۸.۴)، منظور از پیشین، ضرایب سمت راست عبارات مذکور با توجه به u_1 و u_2 است. همین حالت برای پسین برقرار است.	۱.۴
۷۴	اختلاف پسین از پیشین، مخاطره مورد نظر را می‌دهد.	
۷۷	زمان محاسبه شده برای تقریب پسین	۲.۴

فصل ۱

تعاریف و مفاهیم اولیه

۱.۱ مقدمه

تجارب ما در زندگی روزمره ناشی از یک سری تصمیماتی است که مستقیماً روی محیط زیست تاثیرگذار است. پایه تصمیمات ما روی کیفیت زندگی مان اثر می‌گذارد. در حقیقت این تصمیمات بر اساس استنباط یا پیش‌بینی از شرایط موجودی است که ناشی از یک مدل خاص از مشاهدات است. غالباً این مدل‌ها دارای یک روند برجسته و نظم خاص از مشاهدات هستند تا حوادث بعد را پیش‌بینی کنند. گاهی اوقات ساختن و به‌دست آوردن مدل ساده است، اما در اکثر مواقع و در کاربردهای واقعی به مراتب بیش از حد پیچیده است. بنابراین روش خاصی برای طراحی، به علت قابل درک نبودن مدل، در دسترس نیست. در این موارد یک راه حل کلی، استفاده از تقریب‌های ساده برای درک فرایند واقعی که داده‌ها از آن تولید شده‌اند، و استفاده از آن برای استنباط است. در این پایان‌نامه، هدف تحلیل بیزی مدل‌های تصادفی پیچیده‌ای است که دسترسی به تابع درست‌نمایی آن‌ها بسیار دشوار یا ناممکن است. در این صورت، توزیع پسین که مبنای همه استنباط آماری در دیدگاه بیزی است، قابل محاسبه نخواهد بود. با این مقدمه، ما تمرکز خود را روی تحلیل این نوع مدل‌ها، در یک چارچوب بیزی، قرار داده‌ایم. برای برآزش این نوع مدل‌ها، در سال‌های اخیر، رده‌ای از روش‌های محاسباتی معروف به روش‌های محاسبات بیزی تقریبی^۱ (ABC) (بیمونت و همکاران، ۲۰۰۲؛ بلام، ۲۰۱۰؛ فیرنهد و پرانگل، ۲۰۱۰) معرفی شده‌اند. در این پایان‌نامه، به معرفی روش‌های مختلف ABC و استفاده از آن‌ها برای استنباط بیزی مدل‌های تصادفی پیچیده می‌پردازیم. برای ورود به بحث اصلی، در این فصل، ابتدا تعاریف و مفاهیم لازم را ارائه می‌دهیم.

^۱Approximate Bayesian Computing

۲.۱ استنباط آماری

استنباط آماری، یک فرایند استنتاجی در مورد ویژگی‌های یک جامعه، شامل آزمون فرضیه و برآوردیابی است. جامعه شامل همه داده‌های مشاهده شده است. از آنجایی که دسترسی یا تحلیل همه مشاهدات جامعه، به دلیل حجم بالای آن، مشکل است، استنباط آماری بر اساس یک نمونه معقول از جامعه انجام و نتایج حاصل از آن به جامعه مورد نظر تعمیم داده می‌شود. بنابراین، هدف استنباط آماری تعمیم نتایج حاصل از مشاهدات نمونه‌ای به جامعه مورد نظر است. برای این‌که بتوانیم نتیجه‌ای درست از استنباط آماری داشته باشیم، رعایت یک نکته ضروری است، و آن انتخاب مدل واقعی یا مدل نزدیک به فرآیند تولید داده‌ها است. چرا که عمده مشکلات در استنباط آماری به انتخاب مدل نادرست مربوط می‌شود (کونیش و کیت‌گاوا، ۲۰۰۸).

در استنباط آماری مکاتب فلسفی متفاوتی وجود دارند. این مکاتب یا الگوها^۲، منحصر به فرد نیستند و روشی که تحت یک الگو روی مشاهدات به خوبی اجرا می‌گردد غالباً تفسیر متفاوتی نسبت به سایر الگوها روی نمونه مورد نظر دارد. بندی‌پاتای و فورستر (۲۰۱۱) چهار الگو را برای استنباط آماری معرفی می‌کنند:

- آمار کلاسیک یا آمار بسامدی

- آمار بیزی

- آمار مبتنی بر درست‌نمایی

- آمار مبتنی بر معیار اطلاع آکاییک^۳ (AIC).

اما در استنباط آماری بر اساس پارامتر مورد نظر دو دیدگاه وجود دارد: دیدگاه بسامدی، دیدگاه بیزی. در زیر به‌طور خلاصه آن‌ها را شرح می‌دهیم.

۳.۱ دیدگاه بسامدی

در مکتب بسامدی استنباط آماری، پارامتر مورد بررسی ثابت اما نامعلوم است و احتمال به عنوان فراوانی نسبی رخداد پیشامد مورد نظر به ازای تکرار زیاد آزمایش تصادفی، روی مقادیر فضای نمونه و با توجه به پارامتر نامعلوم، تفسیر می‌شود. در این مکتب، معمولاً عملکرد هر فرآیند آماری با میانگین‌گیری روی کل فضای نمونه تعیین می‌شود. این میانگین‌گیری می‌تواند قبل از اجرای آزمایش انجام شود و به داده‌ها بستگی ندارد که باعث شده است به این دیدگاه، از این منظر، نقدهایی وارد شود. به‌طور کلی، دو دیدگاه در استنباط بسامدی وجود دارند:

- فیشر، استنباط آماری را مبتنی بر تابع درست‌نمایی می‌دانست (فیشر، ۱۹۹۲). تابع درست‌نمایی ضابطه مشابهی همچون چگالی توام نمونه دارد، با این تفاوت که، مشاهدات ثابت در نظر گرفته

^۳Akaiik Information Criterion

^۲Paradims

می‌شوند و پارامترها مجاز به تغییر روی همه مقادیر ممکن هستند. او پیچیدگی داده‌ها را با کمک آماره‌های بسنده، کاهش داد به طوری که این آماره‌ها همه اطلاعات موجود در داده‌ها در مورد پارامتر مورد بررسی را دارند (کسلا و برگر، ۲۰۰۲). فیشر همچنین نظریه برآوردیابی درست‌نمایی ماکسیم^۴ (MLE) را ایجاد و توزیع مجانبی برآوردگرهای حاصل از این روش برآوردیابی را معرفی کرد (فیشر، ۱۹۲۲). او کارایی یک برآوردگر را با استفاده از معیاری به نام اطلاع فیشر، اندازه‌گیری کرد، که مقدار اطلاعات در دسترس را در یک مشاهده تکی می‌داد. موضوع‌های مرتبط دیگری که توسط وی در این دیدگاه مطرح شد، عبارتند از: تحلیل واریانس^۵ (گلن، ۲۰۰۵)، تصادفی‌سازی^۶ (فیشر، ۱۹۲۶)، آزمون‌های معناداری^۷ (شام و پورسل، ۲۰۱۴)، آزمون‌های جایگشتی^۸ (پسارین، ۲۰۰۱) و فواصل اتکایی^۹ (هنینگ، ۲۰۰۶؛ بلستات، ۲۰۱۰). وی به عنوان یک دانشمند و آماردان، سهم قابل توجهی در ژنتیک، طرح آزمایش‌ها و استنباط آماری دارد.

• نیمن و پیرسون (۱۹۳۳)، پایه‌های نظریه تصمیم را ایجاد کردند و سپس استنباط آماری را در آن گنجانند. نظریه آن‌ها بر خلاف فیشر اساساً قیاسی^{۱۰} بود. آن‌ها نواحی اطمینان را معرفی و سعی کردند تا راه‌حل‌های بهینه‌ای در رده‌های مجاز^{۱۱} بیابند (نوی و همکاران، ۲۰۰۰). اگر نیاز بود رده‌ها را محدود می‌کردند تا بتوانند راه‌حلی بیابند. برای مثال، در برآوردیابی نقطه‌ای، آن‌ها دریافتند که به‌طور یکنواخت برآوردگر می‌نیمم توان دوم خطا وجود ندارد. بنابراین با محدود کردن رده برآوردگرها به برآوردگرهای ناریب، برآوردگر به‌طور یکنواخت ناریب با کمترین واریانس^{۱۲} (UMVU) را یافتند. والد (۱۹۵۰) این ایده‌ها را با تعریف یک تابع زیان^{۱۳} و سپس تعریف تابع مخاطره^{۱۴} به عنوان امید ریاضی تابع زیان که میانگینی روی فضای نمونه بود، تعمیم داد. سپس قاعده تصمیم ناپذیرفتنی^{۱۵} را تعریف کرد که به‌وسیله همه مقادیر پارامتر تحت پوشش قرار می‌گرفت. با تعریف او، هر قاعده دیگری که ناپذیرفتنی نباشد، پذیرفتنی است. والد از معیارهای بسامدی برای ایجاد چارچوب نظریه تصمیم خود استفاده می‌کرد. اما به‌طور غیرمنتظره‌ای فهمید که رده‌ای از قواعد پذیرفتنی، یک رده از قواعد بیزی را تشکیل می‌دهند. موضوع‌های دیگر در این زمینه شامل: پرتوانترین آزمون‌های یکنواخت در آزمون فرضیه‌ها (والد، ۱۹۴۱)، به‌طور یکنواخت پرتوانترین آزمون‌های ناریب (چوی و همکاران، ۱۹۶۶) و برآوردگرهای جیمز-استاین (استاین، ۱۹۸۱) می‌باشند.

تفاوت دو دیدگاه فیشر و نیمن بنیادی است (ساویج ۱۹۷۶). فیشر به‌شدت از نظر استنتاج با نظریه تصمیم نیمن که استنتاج را با استفاده از منطق استقرایی انکار می‌کرد، مخالفت می‌ورزید. مخالفت وی، به‌ویژه، درباره دیدگاه نیمن-پیرسون شامل موارد زیر بود:

^{۱۰} Deductive

^{۱۱} Allowed classes

^{۱۲} Uniformly Minimum Variance Unbiased

^{۱۳} Loss function

^{۱۴} Risk function

^{۱۵} Inadmissible decision rule

^۴ Maximum Likelihood Estimation

^۵ Analysis of variance

^۶ Randomization

^۷ Significance tests

^۸ Permutation tests

^۹ Fiducial intervals

- برآوردگرهای نارایب تحت تبدیلات پارامتری (بازپارامتری کردن^{۱۶}) یک به یک، بسته نیستند. به این معنی که، اگر برآوردگری مانند T برای پارامتری مانند θ نارایب باشد، $g(T)$ لزوماً برای $g(\theta)$ ، که در آن $g(\cdot)$ یک به یک است، نارایب نیست.
- برآوردگرهای نارایب با اصل درست‌نمایی سازگار نیستند.
- برآوردگرهای نارایب کارا نیستند. فیشر با از بین رفتن اطلاعات که معادل با دور انداختن داده‌ها بود، به شدت مخالف بود.

با این حال، آن‌چه که در حال حاضر به‌عنوان مجموعه ابزار و تکنیک‌های دیدگاه بسامدی استنباط آمار پارامتری شناخته می‌شود، ترکیبی از مفاهیم و روش‌های هر دو گروه است و با وجود اختلافاتی که وجود دارند، از هر دو دیدگاه فیشری و نیمن-پیرسونی به‌عنوان آمار بسامدی، استفاده می‌شود. شاید دلیل این ترکیب آن است که برای حالت‌های مهمی مانند مدل‌های نرمال و دوجمله‌ای، MLE و UMVUE همزمان یکی هستند. دیدگاه بسامدی در مواردی مانند رده خانواده توزیع‌های نمایی، که آماره بسنده می‌نیمال وجود دارد، به خوبی کار می‌کند.

استنباط مبتنی بر تابع درست‌نمایی با استفاده از ایده فیشر اساساً سازنده است. به این معنی که الگوریتم‌ها می‌توانند ساختاری برای حل مسایل بیابند. افرون (۱۹۸۶) به MLE به‌عنوان «جک نایف اصلی» اشاره کرد؛ زیرا ابزاری است که می‌تواند در بسیاری از موقعیت‌ها استفاده گردد. این نوع استنباط ویژگی‌های خوب قابل اتکایی دارد. از جمله این‌که برآوردگرهای درست‌نمایی ماکسیمم، تحت تبدیلات پارامتری پایا هستند؛ با اصل درست‌نمایی سازگار هستند؛ استنباط بسامدی مبتنی بر تابع درست‌نمایی با استنباط بیزی، البته با تفاوت‌هایی، شباهت دارد.

۴.۱ دیدگاه بیزی

مکتب بیزی استنباط آماری مبتنی بر قضیه‌ای شکل گرفت که برای اولین بار توسط یک کشیش به نام توماس بیز در سال ۱۷۶۳ مطرح و پس از مرگ او توسط دوستش، ریچارد پرایس، در مجله‌ای علمی در انجمن سلطنتی انگلستان به چاپ رسید (کسلا و برگر، ۲۰۰۴). قضیه بیز که خیلی هوشمندانه است به وسیله یک احتمال شرطی بیان می‌شود و روشی را برای به‌روز کردن احتمال رخداد‌های مشاهده‌نشده که در ارتباط با رویداد دیگری رخ داده‌اند، ارائه می‌دهد. یعنی با توجه به وقوع رخداد مرتبط، یک احتمال پیشین برای رخداد مشاهده‌نشده داریم، سپس با جمع‌آوری اطلاعات جدید احتمال پسین را به‌روز می‌کنیم. در آمار بیزی، قضیه بیز به‌عنوان اساس استنباط درباره پارامتر نامعلوم توزیع آماری استفاده می‌شود. اجزای کلیدی این رویکرد را می‌توان به صورت زیر بیان کرد:

- از آنجایی که مقدار واقعی پارامتر نامعلوم است، در آمار بیزی فرض می‌شود این مقدار واقعی تحقیقی از یک توزیع احتمالی است. به عبارتی پارامتر یک متغیر تصادفی محسوب می‌شود. این

^{۱۶}Reparametrization

تضادی است که با ایده بسامدی که در آن پارامتر نامعلوم ثابت فرض می‌شود، وجود دارد. قضیه بیز الگوریتم به‌روز کردن استنباط درباره پارامتر واقعی با توجه به اطلاعات مشاهده‌شده است. بنابراین توزیع احتمال پیشینی داریم که توسط آن شانس مشاهده هر مقدار ممکن پارامتر قبل از مشاهده داده‌ها، به گونه‌ای که موجه باشد، محاسبه می‌شود. توزیع احتمال پیشین می‌تواند ذهنی باشد، زیرا هر کس دیگری می‌تواند اعتقاد پیشینی در مورد پارامتر نامعلوم داشته باشد.

- هر گزاره احتمالی درباره پارامتر باید به عنوان «درجه باور» فرد درباره آن پارامتر تفسیر گردد.

- از قاعده احتمالی به‌طور مستقیم استفاده می‌کنیم تا استنباط درباره پارامتر را با توجه به داده‌های مشاهده‌شده انجام دهیم. قضیه بیز توزیع پسینی را نتیجه می‌دهد که بررسی هر مقدار ممکن پارامتر بعد از مشاهده داده‌ها موجه باشد.

- قضیه بیز را با دو منبع اطلاعاتی درباره مقدار پارامتر نامعلوم ترکیب می‌کنیم: اطلاعات پیشین و داده‌های مشاهده‌شده. اطلاعات پیشین، در قالب یک توزیع پیشین، به عنوان وزن و باور نسبی از مقدار واقعی پارامتر قبل از مشاهده داده‌ها تعیین می‌شود. اطلاعات موجود در داده‌های مشاهده‌شده نیز توسط تابع درستنمایی خلاصه می‌شود. قضیه بیز این دو منبع را ترکیب می‌کند و توزیع پسین را که وزن نسبی باورمان از مقدار پارامتر بعد از مشاهده داده‌ها می‌باشد را ارائه می‌دهد.

با توجه به داده‌هایی که در اختیار داریم، قضیه بیز تنها روش سازگار با تغییر باورمان در مورد پارامتر است. استنباط بیزی فقط وابسته به داده‌هایی است که مشاهده شده‌اند، نه آن‌هایی که می‌توانند رخ بدهند اما رخ نداده‌اند. بنابراین استنباط بیزی با اصل درستنمایی سازگار است (برگر و ولپرت، ۱۹۹۸). یک مزیت عمده استنباط بیزی این است که پسین همیشه با یک روش واحد و خودکار مشخص می‌شود: اطلاعات اولیه درباره پارامتر را از چگالی پیشین و اطلاعات موجود در داده‌های مشاهده‌شده را توسط تابع درستنمایی با هم ترکیب می‌کند و چگالی پسین را می‌سازد. بار دیگر تاکید می‌کنیم در دیدگاه بیزی همه استنباط‌ها مبتنی بر توزیع پسین به دست می‌آیند.

۱.۴.۱ توزیع پسین

فرض کنید بردار مشاهدات $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ دارای تابع درستنمایی $\pi(\mathbf{y}|\theta)$ باشد به طوری که $\theta = (\theta_1, \dots, \theta_p) \in \Theta \subset \mathbb{R}^p$. همچنین فرض کنید بردار پارامتر θ دارای تابع چگالی پیشین $\pi(\theta)$ باشد. در این صورت، بنابر قضیه بیز، تابع چگالی توزیع پسین به شکل

$$\pi(\theta|\mathbf{y}) = c\pi(\theta)\pi(\mathbf{y}|\theta) \propto \pi(\theta)\pi(\mathbf{y}|\theta) \quad (1.1)$$

قابل محاسبه است، که در آن

$$c = \frac{1}{\int_{\mathbb{R}^p} \pi(\theta)\pi(\mathbf{y}|\theta)d\theta} \quad (2.1)$$

و به آن ثابت نرمال‌ساز^{۱۷} می‌گویند.

شایان ذکر است این تابع چگالی پسین را زمانی می‌توان محاسبه کرد که تابع درستنمایی قابل محاسبه باشد. در غیر این صورت نمی‌توان از آن برای استنباط بیزی استفاده نمود. انتگرال (۲.۱) به جز در موارد خاص^{۱۸}، صورت بسته‌ای ندارد. بنابراین باید آن را به صورت عددی محاسبه کرد. این محاسبه ممکن است بسیار مشکل باشد، به‌ویژه زمانی که تعداد پارامترها زیاد است.

به‌دست آوردن توزیع پسین از دیدگاه نظری، با استفاده از قاعده بیز، آسان است. یعنی با داشتن توابع درستنمایی و چگالی پیشین و با استفاده از (۱.۱) می‌توان آن را محاسبه کرد. اما در عمل، محاسبه ضریب نرمال‌ساز c می‌تواند بسیار سخت یا حتی غیرممکن باشد. این سختی با افزایش بعد پارامتر شدیدتر هم می‌شود.

بنابراین، یکی از مهم‌ترین چالش‌ها در استنباط بیزی محاسبه تابع چگالی پسین است. به‌ویژه آن‌که معمولاً نمی‌توان فرض کرد که مشاهدات مصداقی از یک نمونه تصادفی از توزیع نرمال باشند. برای برخورد با این چالش، روش‌های عددی و نمونه‌گیری مختلفی معرفی شده‌اند (گلفاند و اسمیت، ۱۹۹۰؛ رابرت و کسلا، ۲۰۰۲؛ تیرنی و همکاران، ۱۹۸۶). در ادامه به برخی از روش‌های نمونه‌گیری که در این پایان‌نامه استفاده می‌شوند، اشاره می‌کنیم.

۵.۱ روش‌های نمونه‌گیری

در این بخش سه روش نمونه‌گیری معمول برای تقریب توزیع پسین را تشریح می‌کنیم.

۱.۵.۱ نمونه‌گیری رد

نمونه‌گیری رد^{۱۹} (RS) روشی است که برای نمونه‌گیری مستقیم از توزیع هدف طراحی شده است (سایمن، ۲۰۰۷). این روش حتی بدون ثابت نرمال‌ساز قابل اجراست. الگوریتم رد، از توزیعی معروف به توزیع پیشنهادی^{۲۰} (که گاهی توزیع آغازین نیز نامیده می‌شود) تولید نمونه می‌کند و نمونه تولیدشده را بر اساس قاعده‌ای که در ادامه توضیح می‌دهیم، به عنوان نمونه‌ای از توزیع پسین می‌پذیرد.

فرض کنید تابع چگالی پسین (بدون ثابت نرمال‌ساز) $\pi(y|\theta)\pi(\theta)$ و $\pi_0(\theta)$ تابع چگالی پیشنهادی باشد. در این روش باید چگالی پیشنهادی طوری انتخاب شود که پسین را بپوشاند. به این معنی که بتوان برای هر θ ، عدد M را به‌طوری یافت که

$$M\pi_0(\theta) \geq \pi(y|\theta)\pi(\theta). \quad (3.1)$$

الگوریتم رد فقط با مقادیر پیشنهادی که در نمونه نهایی پذیرفته شده‌اند، کار می‌کند. فرآیند تولید N نمونه از توزیع پسین بر اساس الگوریتم نمونه‌گیری رد با تعریف $g(\theta) = \frac{\pi(y|\theta)\pi(\theta)}{M\pi_0(\theta)}$ به صورت زیر است:

^{۱۹}Rejection Sampling

^{۲۰}Proposal distribution

^{۱۷}Normalization constant

^{۱۸}برای خانواده‌های پیشین مزدوج در رده خانواده

توزیع‌های نمایی

جدول ۱.۱: الگوریتم نمونه‌گیری رد

الگوریتم نمونه‌گیری رد
تا پذیرش T نمونه مراحل زیر را تکرار کن
۱- θ را از توزیع $\pi_0(\cdot)$ تولید کن
۲- $g(\theta)$ را محاسبه کن
۳- u را از توزیع یکنواخت استاندارد تولید کن
۴- اگر $u < g(\theta)$ ، θ را بپذیر. در غیر این صورت θ را رد کن
۵- به مرحله ۱ بازگرد

با توجه به الگوریتم رد، θ_i ‌های تولید شده از $\pi_0(\theta)$ که برای آن‌ها $u_i < g_i$ ، به‌عنوان نمونه‌هایی از توزیع پسین $\pi(\theta|y)$ پذیرفته می‌شوند. معمولاً تعداد تکرار مراحل الگوریتم از حجم نمونه مورد نظر، T ، به مراتب بیشتر است، مگر آن‌که توزیع پیشنهادی متناسب با توزیع هدف باشد و این به معنی پذیرش همه پیشنهادها است. در این الگوریتم هرچه تفاوت T و تعداد مراحل اجرا بیشتر باشد، الگوریتم از کارایی کمتری برخوردار است. این تفاوت به انتخاب M وابسته است. بنابراین برای داشتن الگوریتمی کارا انتخاب M خیلی مهم است. این انتخاب در موارد متعددی، به ویژه زمانی که بعد پارامتر بالا است، پیچیده و مشکل می‌باشد. برای پذیرفته شدن ماکسیمم تعداد پیشنهادها می‌توان، توزیع پیشنهادی باید شکلی همانند پسین داشته باشد تا آن را بپوشاند (مارتینو و میگز، ۲۰۰۹).

در نمونه‌گیری رد اگر $M = \sup g(\theta)$ باشد، الگوریتم به خوبی کار نمی‌کند. این حالت زمانی اتفاق می‌افتد که توزیع پیشنهادی دمی سبک‌تر از پسین داشته باشد. همچنین زمانی که دم توزیع پیشنهادی خیلی سنگین‌تر از دم توزیع هدف باشد، استفاده از آن نیز کارا نخواهد بود. برای حل این مشکل معمولاً از لگاریتم چگالی پیشنهادی و هدف استفاده می‌کنند. به عنوان مثال، توزیع t با درجه آزادی کوچک برای یک توزیع پیشنهادی دم سنگین در خیلی از کاربردها استفاده می‌شود (لی و کیفوش، ۲۰۱۳؛ کیندرمن و موناهان، ۱۹۷۷).

در این روش اگر توزیع پیشین $\pi(\theta)$ را به عنوان توزیع پیشنهادی $\pi_0(\theta)$ در نظر بگیریم، آن‌گاه $g(\theta) \propto \pi(y|\theta)$. بنابراین قضاوت در مورد پذیرش یا رد نمونه‌های پیشنهادی توسط تابع درستنمایی صورت می‌گیرد.

به راحتی می‌توان نشان داد که نمونه پذیرفته‌شده با این روش، نمونه‌ای از توزیع پسین دلخواه است. برای این کار، تعریف می‌کنیم

$$p(\theta|accept) = \frac{p(accept|\theta)p(\theta)}{p(accept)}.$$

بنابراین، برای هر تکرار و به‌ازای $p(\theta) = \pi_*(\theta)$ ، احتمال پذیرش برابر است با

$$\begin{aligned} p(\text{accept}) &= \int p(\text{accept}|\theta)p(\theta)d\theta \\ &= \int \frac{\pi(y|\theta)\pi(\theta)}{M\pi_*(\theta)}\pi_*(\theta)d\theta \\ &= \frac{1}{M} \int \pi(y|\theta)\pi(\theta)d\theta \\ &= \frac{1}{M}. \end{aligned} \quad (۴.۱)$$

در نتیجه

$$\begin{aligned} p(\theta|\text{accept}) &= \frac{\frac{\pi(y|\theta)\pi(\theta)}{M\pi_*(\theta)}\pi_*(\theta)}{\frac{1}{M}} \\ &= \pi(y|\theta)\pi(\theta). \end{aligned} \quad (۵.۱)$$

با توجه به احتمال پذیرش، می‌توان دریافت که برای تولید هر نمونه به‌طور متوسط M تکرار لازم است. در نتیجه برای کارایی روش باید M کوچک و شبیه‌سازی از توزیع پیشنهادی ساده باشد.

۲.۵.۱ نمونه‌گیری نقاط مهم

در استنباط بیزی معمولاً آماره‌های خلاصه‌شده‌ای از توزیع پسین، مانند میانگین، واریانس و چندک‌های توزیع، مورد علاقه هستند. معمولاً این آماره‌ها را می‌توان با یک امید ریاضی نمایش داد. فرض کنید محاسبه

$$\mathcal{J} = \int_{\Theta} h(\theta)\pi(\theta|y)d\theta = E_{\pi(\theta|y)}(h(\theta)) \quad (۶.۱)$$

مورد نظر باشد. به عنوان مثال، با $h(\theta) = \theta$ ، (۶.۱) میانگین پسین θ را نتیجه می‌دهد. تابع چگالی پسین $\pi(\theta|y)$ می‌تواند دارای صورت بسته یا به‌طور جزئی بسته باشد. برآوردگر مونت کارلویی کلاسیک (۶.۱) به صورت

$$\hat{\mathcal{J}}_{\pi} \simeq \frac{1}{T} \sum_{i=1}^T h(\theta_i) \quad (۷.۱)$$

معرفی می‌گردد که در آن θ ها نمونه تولیدشده از پسین هستند. بنا بر قانون قوی اعداد بزرگ اگر $T \rightarrow \infty$ ، آنگاه

$$\hat{\mathcal{J}}_{\pi} \xrightarrow{a.s.} \mathcal{J}. \quad (۸.۱)$$

دقت داشته باشید که نمونه‌های تولیدشده از پسین می‌توانند مستقل نباشند، که این مورد را در زیربخش ۳.۵.۱ توضیح می‌دهیم.

شایسته است که اشاره کنیم چنان‌چه در رابطه (۶.۱) $\pi(\theta|y)$ را با تابع $\pi_M(y|\theta)$ که تابع درستنمایی تحت مدل M است، و $h(\theta)$ را با تابع پیشین $\pi_M(\theta)$ ، برای پارامترهای تحت مدل M ، جانشین کنیم، آنگاه \mathcal{J} کمیتی معروف به درستنمایی حاشیه‌ای^{۲۱} یا شهود^{۲۲} خواهد بود که در بحث انتخاب مدل و آزمون فرضیه‌های بیزی یک کمیت کلیدی محسوب می‌شود.

اگر تولید مستقیم نمونه از $\pi(\theta|y)$ سخت باشد، یا نواحی چگال $\pi(\theta|y)$ با نواحی چگال $h(\theta)$ مطابقت نداشته باشند، برآوردگر مونت کارلوی کلاسیک (۷.۱) کارا نخواهد بود. در این موارد یک رهیافت جانشین استفاده از نمونه‌گیری نقاط مهم^{۲۳} (IS) است (دوک و همکاران، ۲۰۰۷). در این روش به جای تولید نمونه از $\pi(\theta|y)$ از توزیع دیگری مانند $q(\theta)$ تولید نمونه انجام می‌شود. در رابطه (۶.۱)، \mathcal{J} را می‌توان به صورت

$$\mathcal{J} = \int_{\Theta} \frac{h(\theta)\pi(\theta|y)}{q(\theta)} q(\theta) d\theta = E_q \left[\frac{h(\theta)\pi(\theta|y)}{q(\theta)} \right] \quad (۹.۱)$$

نوشت، که در آن تابع $q(\theta)$ تابع ابزاری^{۲۴} یا پوش^{۲۵} نامیده می‌شود. صورت نمایش (۹.۱) این امکان را میسر می‌سازد که از چگالی‌های دیگری به جز $\pi(\theta|y)$ برای تولید نمونه بتوان استفاده کرد. در این روش $q(\cdot)$ طوری انتخاب می‌شود که تولید نمونه از آن راحت‌تر باشد و برآوردگری کارا تر نتیجه دهد. به طور نمونه این چگالی طوری انتخاب می‌شود که روی تکیه‌گاه خود $q(\theta) \approx |h(\theta)|\pi(\theta|y)$. در این روش نمونه‌گیری، برای برآورد کردن \mathcal{J} با شکل (۹.۱) با داشتن نمونه‌ای از $q(\cdot)$ ، صورت برآوردگر نقاط مهم به شکل زیر است:

$$\hat{\mathcal{J}}_{IS} = \frac{1}{T} \sum_{i=1}^T \frac{h(\theta_i)\pi(\theta_i|y)}{q(\theta_i)}, \quad (۱۰.۱)$$

که در آن $q(\theta) > 0$. این برآوردگر نیز بنا به قانون قوی اعداد بزرگ با احتمال ۱ به \mathcal{J} همگراست. با انتخاب مناسب توزیع ابزاری، واریانس برآوردگر $\hat{\mathcal{J}}_{IS}$ کوچک‌تر از $\hat{\mathcal{J}}_{\pi}$ خواهد بود. همانطور که اوهیگن (۱۹۸۷) اشاره می‌کند، این روش از دو ویژگی مهم برخوردار است. اول، نه تنها برآوردگر به $\pi(\theta|y)h(\theta)$ وابسته است، بلکه به‌طور کامل به انتخاب توزیع پیشنهادی $q(\theta)$ نیز وابسته است. بنابراین با داشتن یک نمونه واحد $(\theta_1, \dots, \theta_T)$ از توزیع پیشنهادی می‌توان برآوردهای مختلفی را برای یک انتگرال یکسان تحت توزیع‌های نمونه‌ای مختلف به دست آورد. به عبارت دیگر با یک نمونه تولیدشده، می‌توان برآوردهای مختلفی را محاسبه کرد. این وابستگی به اطلاعات (کمکی) نامعقول است و اصل درستنمایی را مختل می‌کند (رابرت و کسلا، ۲۰۰۴؛ برگر و ولپرت، ۱۹۸۸). ویژگی دوم آن است که برخلاف روش مونت کارلوی کلاسیک که در آن پس از محاسبه برآورد، همه مقادیر تولیدشده $(\theta_1, \dots, \theta_T)$ را نادیده می‌گیرد، روش نمونه‌گیری نقاط مهم اطلاعات این نمونه‌ها را برای محاسبه برآوردهای دیگر حفظ می‌کند.

^{۲۴}Instrumental

^{۲۵}Envelop

^{۲۱}Marginal likelihood

^{۲۲}Evidence

^{۲۳}Importance Sampling

الگوریتم نمونه‌گیری نقاط مهم در جدول ۲.۱ آمده است. برای اجرای این الگوریتم ابتدا باید وزن‌های

$$w_i = \frac{\pi(y|\theta_i)\pi(\theta_i)/\pi_0(\theta_i)}{\sum_{i=1}^T \pi(y|\theta_i)\pi(\theta_i)/\pi_0(\theta_i)}, \quad i = 1, \dots, T,$$

محاسبه شوند.

جدول ۲.۱: الگوریتم نمونه‌گیری نقاط مهم

الگوریتم نمونه‌گیری نقاط مهم
برای $i = 1$ تا T تکرار کن
۱- θ_i را از توزیع $\pi_0(\cdot)$ تولید کن
۲- پسین $\pi(y \theta)\pi(\theta)$ را محاسبه کن
۳- به ازای هر θ_i ، توزیع پیشنهادی $\pi_0(\theta)$ را محاسبه کن
۴- وزن‌های w_i را محاسبه کن

خروجی الگوریتم ۲.۱ مقادیر (θ_i, w_i) ، $i = 1, \dots, T$ ، شامل نمونه‌های تولید شده و وزن‌های متناظر با آن‌ها می‌باشد.

۳.۵.۱ نمونه‌گیری مونت کارلوی زنجیر مارکوفی

ایده روش‌های نمونه‌گیری مونت کارلوی زنجیر مارکوفی^{۲۶} (MCMC) اولین بار توسط متروپولیس و همکاران (۱۹۵۳) به‌عنوان روشی برای شبیه‌سازی کارا از سطوح انرژی اتم‌ها در ساختارهای بلوری مطرح شد. به دنبال آن در استنباط آماری، به‌ویژه استنباط بیزی، هستینگز (۱۹۷۰) این ایده را پرورش داد و بر اساس نظریه زنجیرهای مارکوف، ویژگی روش‌های نمونه‌گیری MCMC را به‌طور گسترده تحت مطالعه قرار داد. همان‌طور که در بخش‌های قبلی اشاره شد، در استنباط بیزی، توزیع پسین در بیشتر مواقع دارای شکل بسته نیست. در نتیجه نیاز هست تا آن را تقریب بزنیم. تقریب توزیع پسین برای بعدهای بالا کار دشواری است و اشاره کردیم که روش‌های مبتنی بر نمونه‌گیری، یک رده گسترده برای تقریب این توزیع‌ها می‌باشند. گاهی توزیع پسین آنقدر پیچیده است که تولید مستقیم نمونه از آن مشکل است. این دلیل باعث شد تا روش‌های تولید نمونه به‌طور غیر مستقیم معرفی شوند. یک روش غیر مستقیم برای نمونه‌گیری از $\pi(\theta|y)$ آن است که زنجیر مارکوف بازگشتی^{۲۷} و تحویل‌ناپذیری^{۲۸} بسازیم که توزیع مانای آن همان $\pi(\theta|y)$ باشد. نظریه MCMC، پاسخ به چگونگی ساخت چنین زنجیرهایی است (رابرت و کسلا، ۲۰۰۴؛ گلفاند و اسمیت، ۱۹۹۰؛ جورج و کسلا، ۱۹۹۰؛ گمان و گمان، ۱۹۸۴). فرض کنید چنین زنجیری وجود دارد. بنابراین اگر، در اجرا، زنجیر به اندازه کافی بزرگ باشد، آن‌گاه مقادیر شبیه‌سازی شده از زنجیر را می‌توان به عنوان نمونه‌ای از توزیع پسین در نظر گرفت. نمونه‌هایی که توسط این روش‌ها تولید می‌شوند، وابسته هستند و بنابراین برای استنباط بر اساس چنین نمونه‌ای نمی‌توان از قانون اعداد بزرگ استفاده کرد. اما اگر زنجیر تولید شده ارگودیک^{۲۹} باشد (رزنتال و همکاران، ۱۹۹۵؛

^{۲۸}Irreducible

^{۲۹}Ergodic

^{۲۶}Markov Chain Monte Carlo

^{۲۷}Recurrent

والتر، ۱۹۸۲)، آن‌گاه می‌توان از قضیه ارگودیک برای چنین اهدافی استفاده کرد که مشابه قانون اعداد بزرگ ولی برای نمونه‌های وابسته است. تحت این قضیه وقتی $T \rightarrow \infty$ ، برای هر تابع اندازه‌پذیری^{۳۰} مانند L

$$\hat{\mathcal{J}}_{\pi} = \frac{1}{T} \sum_{i=1}^T L(\theta_i) \xrightarrow{a.s.} E_{\pi(\theta|y)}\{L(\theta)\}. \quad (11.1)$$

در اجرای روش‌های نمونه‌گیری MCMC، مطالب مهمی از جمله مکانیسم تولید تحقق برای زنجیر، تعداد زنجیرها و طول زنجیرها و انتخاب نقطه شروع زنجیر مطرح هستند. برای اطلاعات بیشتر در مورد این مطالب به اسمیت (۱۹۸۴)، گرین (۱۹۹۵)، آسموسن (۲۰۰۷) و ریکی (۲۰۱۰)، مراجعه کنید. روش‌های معمول بر اساس نظریه زنجیر مارکوف، با هسته انتقال^{۳۱} (یا ماتریس انتقال در حالت گسسته) آغاز می‌شوند، تا فرآیند مورد علاقه را مدل‌بندی کنند. سپس توزیع مانا را مشخص کرده تا توزیع حدی زنجیر تعیین شود. روش‌های MCMC مسیر معکوس را طی می‌کنند. به این معنی که بر اساس توزیع مانای معلوم، هسته انتقال مشخص می‌شود. هدف این است تا هسته انتقال زنجیر مارکوف (یا توزیع انتقال مارکوف) که با $k(\cdot|\cdot)$ نشان داده می‌شود، طوری تعیین شود که اگر θ معلوم باشد، آن‌گاه توزیع شرطی حالت بعدی زنجیر، θ' ، که آن را با $k(\theta'|\theta)$ نشان می‌دهیم، نیز معلوم باشد. نظریه اصلی و زیربنایی روش‌های MCMC این است که هر زنجیر بازگشتی و تحویل‌ناپذیر، هسته انتقال واحدی خواهد داشت، و هسته انتقال t مرحله‌ای، وقتی $t \rightarrow \infty$ ، به توزیع مانای خود همگرا می‌گردد. بنابراین برای تولید زنجیر با هسته انتقال k و توزیع مانای π باید شرط $\pi k = \pi$ برقرار باشد. این بدان معنی است که با فرض این‌که $\theta \sim \pi$ ، اگر $\theta' \sim k(\theta'|\theta)$ آن‌گاه $\theta' \sim \pi$. چنین زنجیرهایی در شرط زیر، معروف به شرط تعادل دقیق^{۳۲}، صدق می‌کنند:

$$\pi(\theta|y)k(\theta'|\theta) = \pi(\theta'|\y)k(\theta|\theta'). \quad (12.1)$$

این یعنی حرکت زنجیر نسبت به زمان، چه به جلو و چه به عقب، مشابه است. در نتیجه رفتار بازگشتی زنجیر به خوبی درک می‌شود، و این از ویژگی‌های خوب زنجیر برای هر هسته انتقال MCMC است. بر اساس هسته انتقال، الگوریتم‌های مختلفی برای روش‌های نمونه‌گیری MCMC وجود دارند که می‌توان به الگوریتم‌های نمونه‌گیری متروپولیس-هستینگز (متروپولیس و همکاران، ۱۹۵۳؛ هستینگز، ۱۹۷۰) و الگوریتم نمونه‌گیری گیبز (گمان و گمان، ۱۹۸۴؛ گلفاند و اسمیت، ۱۹۹۰) اشاره کرد. به طور کلی الگوریتم MCMC را می‌توان به صورت جدول ۳.۱ ارایه کرد.

با توجه به جدول ۳.۱، احتمال پذیرش الگوریتم MCMC را با p مشخص کردیم. در این صورت با احتمال p ، θ به عنوان نمونه‌ای از توزیع $\pi(\theta|y)$ پذیرفته می‌شود. سپس می‌توانیم با توجه به خروجی الگوریتم که مجموعه‌ای از θ^i ها ($i = 1, \dots, T$) را تشکیل می‌دهد، انتگرال مونت کارلویی یا توزیع پسین را تقریب بزنیم. برای هر الگوریتم MCMC مقدار p تغییر می‌کند. مثلاً برای الگوریتم متروپولیس-هستینگز

^{۳۲}Detailed balance conditon

^{۳۰}Measurable

^{۳۱}Transition kernel

جدول ۳.۱: الگوریتم نمونه‌گیری MCMC

الگوریتم نمونه‌گیری MCMC	
برای $i = 1$ تا T تکرار کن	
۱-	هسته مارکوف $k(\cdot \cdot)$ با توزیع مانای $\pi(\cdot)$ را ایجاد کن
۲-	زنجر مارکوف $\theta \sim \pi$ را تولید کن
۳-	θ را از هسته مارکوف $k(\theta \theta^{(i-1)})$ تولید کن
۴-	قرار بده $\theta = \begin{cases} \theta^{(i)} & \text{با احتمال } p \\ \theta^{(i-1)} & \text{با احتمال } 1-p \end{cases}$

مقدار احتمال پذیرش برابر است با

$$p(\theta, \theta^{(i-1)}) = \min \left\{ 1, \frac{\pi(\theta|y)}{\pi(\theta^{(i-1)}|y)} \frac{k(\theta^{(i-1)}|\theta)}{k(\theta|\theta^{(i-1)})} \right\}. \quad (13.1)$$

۶.۱ نرم‌های آماری

در جبر خطی، آنالیز تابعی و زمینه‌های دیگر ریاضی، نرم تابعی است که در فضای برداری (پیوسته)، یک فاصله اکید مثبت یا صفر را به هر بردار نسبت می‌دهد. علاوه بر نرم از نیم‌نرم^{۳۳} نیز استفاده می‌گردد، که طول صفر را به بردار غیرصفر نیز نسبت می‌دهد. نرم‌ها دارای ویژگی‌هایی مانند ضرب‌پذیری و جمع‌پذیری هستند. به عنوان مثال، نرم اقلیدسی برای هر بردار برابر طول بردار است و به همین خاطر به عنوان اندازه شناخته می‌شود. همچنین فضای برداری که در آن نرم تعریف شده است، را فضای برداری نرم‌دار می‌گویند.

تعریف ۱.۶.۱. فرض کنید V یک فضای برداری روی میدان F از اعداد مختلط باشد. در این صورت تابع $\rho: V \rightarrow R$ را نرم گویند، هرگاه برای هر $a \in F$ و هر $u, v \in V$ دارای سه ویژگی زیر باشد:

۱. $\rho(av) = |a|\rho(v)$ (خاصیت همگنی مطلق^{۳۴} یا مقیاس‌پذیری مطلق^{۳۵})

۲. $\rho(u+v) \leq \rho(u) + \rho(v)$ (نامساوی مثلث^{۳۶} یا خاصیت زیرجمعی^{۳۷})

۳. اگر $\rho(v) = 0$ ، آن‌گاه $v = 0$ (نقاط مجزا)

^{۳۶} Inequality triangle

^{۳۷} Subadditivity

^{۳۳} Seminorm

^{۳۴} Homogeneity absolute

^{۳۵} Scalability absolute

توجه کنید یک نیم‌نرم ویژگی ۳ را ندارد. دو نرم (یا نیم‌نرم) ρ_1 و ρ_2 روی فضای برداری V معادل هستند اگر دو ثابت مثبت c و C وجود داشته باشند، به طوری که برای هر بردار v در V داشته باشیم

$$c\rho_2(v) \leq \rho_1(v) \leq C\rho_2(v). \quad (14.1)$$

در متون آماری، تابع نرم را با نماد $\|\cdot\|$ نیز نشان می‌دهند.

۱.۶.۱ انواع نرم‌ها

در این قسمت انواع نرم‌های معمول را معرفی می‌کنیم.

نرم قدر مطلق

تابع نرم قدر مطلق، یک نرم در فضای یک‌بعدی برای اعداد حقیقی و مختلط است، که به صورت زیر تعریف می‌شود:

$$\rho(x) = |x|.$$

p - نرم

تابع p -نرم در فضای برداری L^p برای $p \geq 1$ ، که در آن p یک عدد حقیقی است، به صورت زیر تعریف می‌شود:

$$\rho(\mathbf{x}) = \|\mathbf{x}\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{1/p},$$

که در آن x یک بردار n بعدی است. برای p های مشخصی توابع نرم معرفی شده‌اند.

- اگر $p = 1$ آن‌گاه $\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$ که به نرم تاکسی نیز معروف است.

- اگر $p = 2$ آن‌گاه $\|\mathbf{x}\|_2 := \sqrt{x_1^2 + \dots + x_n^2}$ که تعریف نرم اقلیدسی است.

- اگر $p = \infty$ آن‌گاه $\|\mathbf{x}\|_\infty := \max(|x_1|, \dots, |x_n|)$ که به آن نرم ماکسیمم گفته می‌شود.

توابع نرم دیگری نیز وجود دارند که در برخی از زمینه‌ها استفاده می‌شوند. به‌عنوان مثال اگر $\mathbf{x} \in \mathbb{R}^4$ ، آن‌گاه یک نرم معمول عبارتست از $\rho(\mathbf{x}) := 2|x_1| + \sqrt{3|x_2|^2 + \max(|x_3|, 2|x_4|)^2}$. با این وجود در این پایان‌نامه، از نرم‌های اقلیدسی و تاکسی استفاده می‌کنیم.

۷.۱ هسته‌ها

عبارت هسته^{۳۸} در آمار دارای معانی مختلفی است.

^{۳۸}Kernel

- در استنباط بیزی، منظور از هسته یک تابع (چگالی) احتمال، تابعی است که در آن هر عاملی که وابسته به متغیرها نیست، حذف شده است. به عبارتی، هسته همان تابع (چگالی) احتمال (پسین) بدون ثابت نرمال‌ساز تعریف می‌شود.
- در متون یادگیری ماشینی^{۳۹}، منظور از هسته معمولاً تابعی است که ساختار ماتریس واریانس را در مدل‌های چندمتغیره مشخص می‌کند.
- در آمار ناپارامتری، هسته به عنوان یک تابع وزنی در تکنیک‌های برآورد تابع چگالی به کار می‌رود.
- در برآورد تابع رگرسیونی به روش ناپارامتری، برای برآورد امید ریاضی شرطی متغیر پاسخ به شرط متغیرهای تبیینی مورد استفاده قرار می‌گیرد.
- همچنین در سری‌های زمانی به عنوان دوره‌نگار^{۴۰} در برآورد تابع چگالی طیفی^{۴۱} به کار می‌رود.

در این پایان‌نامه، منظور از توابع هسته همان توابع هسته آمار ناپارامتری است. در استفاده از این توابع برای اجرای برآوردیابی ناپارامتری، باید پارامتر تنظیم‌کننده‌ای^{۴۲} را که به پهنای نوار^{۴۳} یا پارامتر هموارساز^{۴۴} معروف است، مشخص کنیم. تغییر این پارامتر بر عملکرد روش خیلی تاثیر دارد و در نتیجه در تعیین آن باید دقت لازم را فرض کرد. یک تابع هسته مانند $\mathcal{K}(\cdot)$ باید در دو شرط زیر نیز صدق کند:

$$\int_{-\infty}^{+\infty} \mathcal{K}(u) du = 1 \quad ۰.۱$$

$$\mathcal{K}(-u) = \mathcal{K}(u), \quad u \text{ برای هر مقدار}$$

شرط دوم یک ویژگی تقارن را به تابع هسته القا می‌کند. اگر $\mathcal{K}(\cdot)$ هسته مورد نظر باشد، آنگاه تابع $\mathcal{K}^*(\cdot)$ را وقتی $\lambda > 0$ باشد می‌توان به صورت $\mathcal{K}^*(x) = \lambda \mathcal{K}(\lambda x)$ تعریف کرد. شایع‌ترین توابع هسته که مورد استفاده قرار می‌گیرند را در جدول ۴.۱ آورده‌ایم. بین دو تابع نرم و هسته، می‌توان رابطه

$$\mathcal{K}_\epsilon(y) = \mathcal{K}\left(\frac{\|y\|}{\epsilon}\right), \quad (۱۵.۱)$$

را تعریف کرد، که در آن ϵ پارامتر پهنای نوار و $\mathcal{K}(\cdot)$ تابع هسته (جدول ۴.۱) است. در این پایان‌نامه، در مثال‌های شبیه‌سازی‌شده، از دو هسته نرمال و یکنواخت استفاده کرده‌ایم.

۸.۱ مدل‌های تصادفی پیچیده

در شاخه‌های مختلف علوم، موقعیت‌های کاربردی مختلفی را می‌توان برشمرد که مدل تصادفی منطقی برای آن‌ها پیچیده و برازش و استنباط در آن‌ها کار دشواری است. پیچیدگی دسته‌ای از این مدل‌ها، به

^{۴۲}Tuning parameter

^{۴۳}Bandwidth

^{۴۴}Smoothing parameter

^{۳۹}Machine learning

^{۴۰}Periodogram

^{۴۱}Spectral density

جدول ۴.۱: توابع هسته آماری رایج

توابع هسته		
	$\mathcal{K}(u) = \frac{1}{2} \mathcal{I}_{ u \leq 1}$	هسته یکنواخت (Uniform)
	$\mathcal{K}(u) = (1 - u) \mathcal{I}_{ u \leq 1}$	هسته مثلثی (Traingular)
	$\mathcal{K}(u) = \frac{3}{4} (1 - u^2) \mathcal{I}_{ u \leq 1}$	هسته اپانچنیکف (Epanechnikov)
	$\mathcal{K}(u) = \frac{15}{8} (1 - u^2)^2 \mathcal{I}_{ u \leq 1}$	هسته درجه چهارم (Quartic)
	$\mathcal{K}(u) = \frac{35}{24} (1 - u^2)^3 \mathcal{I}_{ u \leq 1}$	هسته سه‌وزنی (Triweight)
	$\mathcal{K}(u) = \frac{5}{8} (1 - u ^3)^3 \mathcal{I}_{ u \leq 1}$	هسته سه‌مکعب (Tricube)
	$\mathcal{K}(u) = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}u^2}$	هسته نرمال (Normal)
	$\mathcal{K}(u) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) \mathcal{I}_{ u \leq 1}$	هسته کسینوسی (Cosine)
	$\mathcal{K}(u) = \frac{1}{e^u + 2 + e^{-u}}$	هسته لوجستیک (Logistic)
	$\mathcal{K}(u) = \frac{1}{2} e^{-\frac{ u }{\sqrt{2}}} \sin\left(\frac{ u }{\sqrt{2}}\right) + \frac{\pi}{4}$	هسته سیلورمن (Silverman)

دلیل محاسبه سخت و حتی ناممکن تابع درستنمایی آن‌هاست. این پیچیدگی، استنباط بیزی این رده از مدل‌ها را نیز ناممکن و دشوار می‌سازد.

در این پایان‌نامه به دنبال معرفی راه‌حل ارائه‌شده برای این مشکل هستیم که در سال‌های اخیر مورد توجه قرار گرفته است. اما قبل از آن، برای درک بهتر پیچیدگی‌های مورد اشاره، چند مدل تصادفی با این ویژگی را ارائه می‌دهیم. در این مثال‌ها سعی کرده‌ایم تا حد امکان نمونه‌های ساده و قابل درک را برای خواننده غیرمتخصص بگنجانیم. البته دقت کنید که در عین سادگی مدل، محاسبه تابع درستنمایی آن‌ها مشکلی اساسی است.

۱.۸.۱ مدل $k - g$

توزیع $k - g$ بخشی از خانواده توزیع‌های چندکی^{۴۵} و توابعی از چندک نرمال استاندارد، $\Phi(y)$ ، هستند. این توزیع، انعطاف‌پذیر است و برای داده‌های مدل غیر استاندارد به‌کار می‌رود (هینس، ۱۹۹۸). آن با یک تابع توزیع معکوس به‌صورت زیر تعریف می‌شود:

$$Q(y, A, B, c, g, k) = A + B \left(1 + c \frac{1 - e^{-g\Phi(y)}}{1 + e^{-g\Phi(y)}} \right) (1 + \Phi(y)^2)^k \Phi(y), \quad (16.1)$$

که در آن $\Phi(y)$ ، چندک نرمال استاندارد است. همچنین A و B به ترتیب پارامترهای مکان و مقیاس، و g و k به ترتیب چولگی^{۴۶} و کشیدگی^{۴۷} هستند. پارامتر c را معمولاً ثابت و برابر 0.8 در نظر می‌گیرند. در بین این پارامترها فقط $B > 0$ و $k > -1/2$ محدودیت دارند (راینر و مک‌گالیوری، ۲۰۰۲). شکل تابع درستنمایی این توزیع در دسترس نیست و برای این مدل تابع درستنمایی به‌سختی و با استفاده از روش‌های عددی محاسبه می‌گردد. مک‌گالیوری و همکاران (۲۰۰۲) تابع درستنمایی را برای n نمونه تصادفی از توزیع $k - g$ به صورت

$$\begin{aligned} \pi(\mathbf{y}|\theta) &= \prod_{i=1}^n f_Y(y_i|\theta) \\ &= \prod_{i=1}^n \frac{\partial}{\partial y_i} Q_Y^{-1}(y_i|\theta) \\ &= \prod_{i=1}^n Q'_Y(Q_Y^{-1}(y_i|\theta)|\theta)^{-1} \end{aligned} \quad (17.1)$$

نشان دادند، که در آن

$$Q'_Y(y|\theta) = \frac{\partial Q_Y}{\partial y} = \frac{\partial Q_Y}{\partial \Phi^{-1}(y)} \frac{\partial \Phi^{-1}(y)}{\partial y}$$

و $\frac{\partial \Phi^{-1}(y)}{\partial y} = \sqrt{2\pi} \exp((\Phi^{-1}(y))^2/2)$ است و این‌که در آن $Q_Y^{-1}(y_i|\theta)$ شکل بسته‌ای ندارد. اما شبیه‌سازی از مدل $k - g$ بسیار ساده و سرراست است. این شبیه‌سازی را با استفاده از روش ABC انجام می‌دهند. آلینگ‌هام و همکاران (۲۰۰۹) از روش ABC برای تحلیل این مدل با استفاده از داده‌های شبیه‌سازی‌شده، پرداختند.

^{۴۷}Kurtosis

^{۴۵}Quantile distribution

^{۴۶}Skewness

۲.۸.۱ مدل‌های α -پایدار

توزیع α -پایدار از تابع مشخصه به شکل زیر تشکیل شده است که در آن تابع مشخصه از n متغیر مستقل و هم‌توزیع (Y_1, \dots, Y_n) دارای توزیع α -پایدار به دست می‌آید:

$$\phi_Y(t) = \begin{cases} \exp(i\delta t - \gamma^\alpha |t|^\alpha [1 + i\beta \tan \frac{\pi\alpha}{2} \operatorname{sgn}(t)(|\gamma t|^{1-\alpha} - 1)]) & \alpha \neq 1 \\ \exp(i\delta t - \gamma |t| [1 + i\beta \frac{\pi}{2} \operatorname{sgn}(t)(\ln(\gamma |t|))]) & \alpha = 1 \end{cases}$$

که در آن $\operatorname{sgn}(t) = t/|t|$ و $i^2 = -1$ است (ساموروفسکی و تاکیو، ۱۹۸۴). همچنین $\alpha \in (0, 2)$ ، نرخ تباهی دم^{۴۸}، $\beta \in [-1, 1]$ درجه و نشان عدم تقارن^{۴۹} (چولگی) است. $\gamma > 0$ پارامتر مقیاس، و $\delta \in \mathbb{R}$ پارامتر مکان است (لوی، ۱۹۲۴). پارامتر α توان مشخصه نامیده می‌شود، که با کوچکی یا بزرگی آن، دم سنگین یا سبک خواهیم داشت. توزیع‌های $N(\alpha = 2, \beta = 0)$ و کوشی $c(\alpha = 1, \beta = 0)$ زیرمجموعه‌های رام‌نشده این خانواده را تشکیل می‌دهند. با توجه به توزیع α -پایدار، تابع درست‌نمایی آن شکل مشخصی ندارد. حتی محاسبه آن به روش‌های عددی بسیار مشکل است. از این رو نیاز به روشی برای حل این مساله احساس می‌شود. پیترو و سیسون (۲۰۱۰) از روش ABC برای حل این مساله استفاده کردند.

۳.۸.۱ چگالی توام میانه و انحراف مطلق میانه

فرض کنیم که نمونه تصادفی مستقل و هم‌توزیع (y_1, \dots, y_n) را داریم. در این صورت آماره انحراف مطلق میانه^{۵۰} (mad) به صورت

$$\operatorname{mad}(\mathbf{y}) = \operatorname{med}(|y_i - \operatorname{med}(\mathbf{y})|) \quad i = 1, \dots, n$$

تعریف می‌شود. حال فرض کنیم که $y_i \sim N(0, \sigma^2)$. در این صورت آماره دو-بعدی

$$S(\mathbf{y}) = (\operatorname{med}(\mathbf{y}), \operatorname{mad}(\mathbf{y}))$$

که بسنده نیست، چگالی توام آن شکل بسته‌ای ندارد و بدیهی است که تابع درست‌نمایی توام آن نیز شکل بسته نخواهد داشت.

۹.۱ روش‌های درست‌نمایی آزاد

در این زیربخش به طور خلاصه، به معرفی رده‌ای از روش‌ها برای مدل‌های آماری بیزی می‌پردازیم که شامل توابع درست‌نمایی رام‌نشده^{۵۱} هستند. این رده از روش‌ها معروف به درست‌نمایی آزاد^{۵۲} هستند. بخش عمده‌ای از این رده به محاسبات بیزی تقریبی (ABC) شناخته می‌شوند. عبارت رام‌نشده در

^{۵۱}Intractable

^{۵۲}Free likelihood

^{۴۸}Tail decay

^{۴۹}Asymmetry

^{۵۰}Median absolute deviation

تابع درست‌نمایی به معنی بسته نبودن صورت تابع درست‌نمایی است؛ به این معنی که محاسبه آن به طور مستقیم غیرممکن است یا هزینه بالایی به دنبال خواهد داشت. روش‌های ABC با انگیزه مدل‌بندی بیزی مدل‌ها با توابع درست‌نمایی رام‌نشده معرفی شده‌اند (مارین و همکاران، ۲۰۱۲). در این روش‌ها به جای توزیع پسین واقعی $\pi(\theta|y)$ از نسخه تقریبی $\pi_{ABC}(\theta|y)$ برای اهداف استنباط استفاده می‌شود. در ادامه نحوه محاسبه $\pi_{ABC}(\theta|y)$ را توضیح خواهیم داد. استفاده از روش‌های ABC نیازمند چندین مولفه است:

- توزیع پسین رام‌نشده، که با شبیه‌سازی تقریب زده می‌شود.
- تکنیک شبیه‌سازی داده‌ها از مدل رام‌نشده.
- داشتن مجموعه‌ای از پارامترها کمیت‌های (S, ρ, ϵ) که در آن
- S : آماره خلاصه‌شده^{۵۳} برای داده‌های واقعی و شبیه‌سازی شده
- ρ : معیاری برای اندازه‌گیری تفاوت بین دو بردار مشاهدات یا آماره خلاصه‌شده استفاده می‌شود.
- ϵ : یک سطح تحمل یا دقتی است که برای سنجش نزدیکی دو بردار یا آماره (بر حسب تابع نرم) استفاده می‌گردد.

در روش درست‌نمایی آزاد شبیه‌سازی از توزیع پسین صورت می‌گیرد که با قرار دادن یک جمله اضافی در پسین مدل انجام می‌گیرد. این توزیع پسین به صورت زیر محاسبه می‌شود:

$$\pi(\theta, x|y) \propto \pi(y|x, \theta)\pi(x|\theta)\pi(\theta), \quad (18.1)$$

که در آن $x \in D$ یک بردار از مشاهدات شبیه‌سازی شده است، که روی فضای مشابهی از مشاهدات واقعی $y \in D$ ، بر اساس درست‌نمایی $x \sim \pi(x|\theta)$ تولید می‌گردند. تابع $\pi(y|x, \theta)$ به عنوان وزن پسین رام‌نشده تعریف می‌شود. برای x هایی که آماره خلاصه‌شده آن‌ها به $S(y)$ نزدیک باشند، $S(x) \approx S(y)$ ، وزن پسین بزرگ و برای آن‌هایی که اختلاف بین $S(x)$ و $S(y)$ زیاد باشد، این وزن کم است. زمانی که $S(x) = S(y)$ ، وزن پسین ماکسیمم می‌شود.

در نتیجه توزیع هدف به دست آمده به عنوان توزیع حاشیه‌ای (۱۸.۱) به صورت زیر بیان می‌شود:

$$\begin{aligned} \pi_{ABC}(\theta|y) &\propto \int_{\mathcal{X}} \pi(y|x, \theta)\pi(x|\theta)\pi(\theta)dx \\ &\propto \pi(\theta) \int_{\mathcal{X}} \pi(y|x, \theta)\pi(x|\theta)dx \\ &= \pi(\theta) E_{\pi(x|\theta)}[\pi(y|x, \theta)] \\ &\approx \frac{\pi(\theta)}{s} \sum_{i=1}^s \pi(y|x^i, \theta). \end{aligned} \quad (19.1)$$

با توجه به عبارت (۲۰۱)، $x^1, \dots, x^s \sim \pi(x|\theta)$ ، تقریبی از s مجموعه داده ساختگی (شبهه سازی) از تابع درست‌نمایی (رام‌نشدنی) می‌باشند. در واقع نمونه‌گیری به طور مستقیم از $\pi_{ABC}(\theta|y)$ به وسیله تقریب انتگرال (۱۹۰۱) با انتگرال مونت کارلو در قبال محاسبه هر پسین $\pi_{ABC}(\theta|y)$ انجام می‌شود. امید عبارت (۱۹۰۱)، اولین بار توسط مجرم و همکاران (۲۰۰۳) مطرح شد، و از آن به بعد توسط چند نویسنده دیگر به کار گرفته شد (پیتر و همکاران (۲۰۰۸)، ریوس و پتیت (۲۰۰۵)، راتمن (۲۰۰۹)، تونی (۲۰۰۹)).

در ادامه به اختصار در مورد هرکدام از مولفه‌های تقریب پسین ABC توضیحاتی را ارائه می‌دهیم.

۱۰.۱ تولید تحققی‌هایی از مدل

روش درست‌نمایی آزاد شامل یک فرضیه کلیدی است که در اجرا مورد نیاز می‌باشد، که آن قابلیت تولید تحققی‌های موثر از مدل است. در نتیجه داده‌های ساختگی را تولید می‌کند تا جایگزینی برای محاسبه نقطه به نقطه درست‌نمایی گردد. با توجه به روش درست‌نمایی آزاد که در عبارت (۱۸۰۱) بیان شد و شامل توزیع پسین نشانیده بود، واضح است که درست‌نمایی آزاد حاشیه‌ای مرتبط با پارامتر هدف، تقریبی از توزیع پسین بود که در عبارت (۱۹۰۱) بیان شد، و آن تحققی‌ها، شبهه سازی از s مجموعه داده بود که عموماً با $x^{(s)}$ مشخص شده بودند.

امید این فرضیه در مقاله پیتر و همکاران (۲۰۰۹) ارائه شد، که در آن شبهه سازی از مدل نمی‌تواند به صورت مستقیم انجام شود. برای رفع این مشکل از فرایند بوت استرپ خاصی در روش درست‌نمایی آزاد استفاده می‌کنند تا داده‌های ساختگی x تولید گردند.

۱.۱۰.۱ آماره‌های خلاصه شده

در روش درست‌نمایی آزاد، مجموعه‌ای از داده‌های ساختگی x تولید می‌شود، و در ادامه به محاسبه خلاصه داده‌های واقعی و ساختگی می‌پردازیم، که به ترتیب با $S(y)$ و $S(x)$ بیان شده است. انتخاب‌های زیادی برای آماره‌های خلاصه شده وجود دارد. در ابتدا روش ABC، $S(y) = y$ را انتخاب می‌کند، که به طور مستقیم به مقایسه داده‌های واقعی و شبهه‌سازی شده می‌پردازد. خیلی زود پی برده شد که برای مشاهدات روی دامنه پیوسته، این مقایسه بیش از حد محدود است. در حالت ایده‌آل، دلخواه این است که برداری از آماره‌های خلاصه شده S را که شامل آماره‌های بسنده هستند، داشته باشیم. بنابراین با توجه به قضیه فاکتورگیری نیمن-فیشر، این انتخاب از خلاصه‌ها همه اطلاعات مورد نظر در مدل را در بر می‌گیرد. سپس طبق این انتخاب، روش ABC به مقایسه آماره‌های بسنده روی داده‌های واقعی و شبهه‌سازی، بدون ارزیابی، می‌پردازد. در اکثر اجرای مدل درست‌نمایی آزاد، آماره‌های بسنده به دست نمی‌آید. در حالت کلی برداری از آماره‌های خلاصه شده S عموماً مقداری درجه نزدیک-بسندگی را برای θ از طریق $\pi(x|\theta)$ ایجاد می‌کنند. چندین آماره بسنده دیگر برای آماره خلاصه شده S پیشنهاد شده است (برای مثال پیتر و همکاران (۲۰۰۹) را ببینید). نمونه‌های مشهور شامل مجموع‌های تجربی قابل اجرا همچون: گشتاورها، چندک‌ها و صدک‌ها، برآورد تابع مشخصه تجربی، آماره‌های متعامد و

تبدیل پارامترهای مدل θ به ϕ برای برآوردگرهایی که معلوم هستند. توجه شود که در ABC، می‌توان داده‌ها را از طریق پارامترهای تبدیل مقایسه کرد، تا زمانی که از مدل تبدیل به دست آیند. در چندین مثالی که در پیترو و همکاران (۲۰۰۹) نشان داده شد، این واقعیت از نظر کارایی و تعمیم روش بسیار مفید است. اخیراً تعمیم بهتری توسط راتمن و همکاران (۲۰۰۹) ارائه شد، که کاربرد دیگری برای S مورد بررسی قرار دادند.

۲.۱۰.۱ تابع فاصله (نرم)

در روش درست‌نمایی آزاد با به دست آمدن آماره خلاصه شده $S(x)$ و $S(y)$ ، فاصله متریک بین این آماره خلاصه شده را عموماً با $\rho(S(x), S(y))$ نشان می‌دهند.

توجه داشته باشید که به‌ویژه معیارهای فاصله‌ای که حاوی اطلاعات مربوط به همبستگی بین آماره خلاصه شده هستند، برآوردی از پسین حاشیه‌ای $\pi_{ABC}(\theta|y)$ می‌دهند که بر اساس هزینه محاسباتی محدود به مراتب خیلی دقیق‌تر و کارایی بیشتری در الگوریتم‌های شبیه‌سازی مورد استفاده دارند.

به صورت تئوری شکل فاصله ρ تا زمانی که شرط زیر برقرار باشد مهم نیست. یعنی

$$\rho(S(x), S(y)) = 0 \quad \text{اگر و فقط اگر} \quad S(x) = S(y),$$

این شرط تضمین می‌کند که توزیع حدی درست است. این را می‌توان با استفاده از تحمل مقایسه‌ای ϵ تعمیم داد، که به مقایسه داده‌های واقعی و شبیه‌سازی شده در آماره خلاصه شده می‌پردازد، به طوری که وقتی سطح تحمل $\epsilon \rightarrow 0$ ، در نتیجه

$$\lim_{\epsilon \rightarrow 0} \pi(\theta) \pi(\rho(S(x), S(y)) \leq \epsilon) = \pi(\theta|S(y))$$

به خصوص زمانی که تحمل ϵ کاهش می‌یابد، هزینه‌های محاسباتی افزایش می‌یابند. برای مجموعه‌ای از آماره‌های خلاصه شده، افزایش در محاسبات وابسته به انتخاب تابع فاصله است، که این نیز بر دقت برآورد $\pi_{ABC}(\theta|y)$ تاثیرگذار است. از این رو باید یک تابع فاصله‌ای مناسبی به کار ببریم تا برای تحمل به اندازه کافی بزرگ ϵ تقریب خوبی برای $\pi(\theta|S(y)) = \pi(\theta|\rho(S(y), S(x)) \leq \epsilon)$ ایجاد شود و در نتیجه هزینه محاسباتی کاهش یابد. بحث در مورد این که چه ϵ انتخاب کنیم در چندین مقاله همچون پیترو و همکاران (۲۰۰۹) و نوات و همکاران (۲۰۰۹) ارائه شده است.

۳.۱۰.۱ تصمیم و تابع وزنی

با داشتن ϵ که فاصله بین آماره‌های خلاصه شده و شبیه‌سازی شده است، گام بعدی وزن دادن به خصوص به $\pi(y|x^s, \theta)$ در عبارت (۱۹.۱) است. چندین انتخاب برای تابع وزنی وجود دارد. در بیشتر مقالات از یک تابع وزنی استفاده شده که به طور مستقیم خلاصه داده‌های واقعی و شبیه سازی شده را مقایسه می‌کند، یعنی

$$\pi(y|x, \theta) \propto \begin{cases} 1 & \text{اگر } \rho(S(y), S(x)) = 0 \\ 0 & \text{در غیر این صورت} \end{cases}$$

بنابراین تصمیم در مورد به‌کار بردن آن مفهومی را تحت عنوان نزدیک بسندگی طبق یک سطح تحمل مشخص مطرح می‌کند، که به طور ایده‌آل، آماره‌های بسنده درجه تقریب ABC را به طور مستقیم کنترل می‌کنند. در ادامه تابع وزنی دیگری موسم به تابع وزنی تصمیم نه چندان سخت بیان شده است.

$$\pi(y|x, \theta) \propto \begin{cases} 1 & \text{اگر } \rho(S(y), S(x)) \leq \epsilon \\ 0 & \text{در غیر این صورت} \end{cases}$$

این تابع وزنی هنوز هم مورد استفاده است، اما در چندین مقاله همچون مقاله مارین و همکاران (۲۰۱۲) اشاره شده است که به دلیل افزایش نرخ رد این تابع وزنی ناکارا می‌باشد. شکل‌های دیگری از تابع وزنی وجود دارد همچون هسته اپانچنیکف و هسته گوسی که در بخش قبل به آن اشاره کردیم.

۴.۱۰.۱ سطح تحمل

انتخاب تحمل فرض مهمی است. هنگامی که تاثیر دقت در برآورد تقریب $\pi_{ABC}(\theta|y)$ مطرح است، انتخاب‌های فراوانی وجود دارند که عموماً وابسته به الگوریتم‌های استفاده شده در نمونه‌گیری از پسین هستند. برای مثال الگوریتم رد از یک تحمل ثابت استفاده می‌کند. نسخه‌های پیچیده‌تر همچون زنجیر مارکوف مونت کارلویی و نمونه‌گیری نقاط مهم متمایل به استفاده از تحمل ملایم شده یا نوردیده هستند. اخیراً کار روی نوردیدن طرح پویا و سازوار انجام می‌شود. (جزئیات بیشتر را در مقاله پیتز و همکاران (۲۰۰۹) و دل مورل و همکاران (۲۰۰۸) ببینید)

فصل ۲

محاسبات بیزی تقریبی

۱.۲ مقدمه

ایده روش محاسبات بیزی تقریبی (که به اختصار با ABC نشان دادیم) اولین بار توسط روبین (۱۹۸۴) مطرح شد. او به توصیف مکانیسم یک روش نمونه‌گیری مبتنی بر شبیه‌سازی پرداخت که حاصل آن تولید نمونه‌ای از توزیع پسین است. در واقع روبین برای مدل‌هایی که به نسبت رام‌نشدنی^۱ بودند، راه‌حلی تقریباً خودکار بر اساس شبیه‌سازی ارائه کرد. البته شایسته است اشاره کنیم که فرانسویس گالتون در سال ۱۸۰۰ یک اجرای فیزیکی از یک طرح نمونه‌گیری ABC با یک پارامتر نامعلوم و یک تک‌مشاهده را ارائه داد.

نکته اساسی و جالبی که توسط روبین مطرح شد این بود که در استنباط بیزی، آماردانان نباید فقط به مدل‌های تحلیلی رام‌شدنی و ساده تمرکز کنند و در مقابل به دنبال معرفی روش‌های محاسباتی باشند که برآوردی از توزیع پسین پارامتر مورد علاقه را نتیجه می‌دهد. به این ترتیب، طیف وسیع‌تری از مدل‌های پیچیده‌تر اما واقعی‌تر بررسی می‌گردد.

به نظر می‌رسد دیگ و گریتون (۱۹۸۴) اولین فرآیند شبیه‌سازی برای انجام استنباط آماری، در مواردی که تابع درست‌نمایی رام‌نشدنی هست، را ایجاد کردند. آن‌ها استفاده از یک طرح شبیه‌سازی متقارن را برای تقریب تابع درست‌نمایی موقعی که محاسبه آن دشوار یا نشدنی است، پیشنهاد کردند. روش پیشنهادی آن‌ها مبتنی بر تعریف یک شبکه^۲ در فضای پارامتر است تا با استفاده از آن، تابع درست‌نمایی را با اجرای چندین شبیه‌سازی برای هر نقطه از شبکه و استفاده از روش‌های هموارسازی^۳، تقریب بزند. اگرچه روش دیگ و گریتون مرز جدیدی را برای مدل‌بندی فرآیندهای تصادفی پیچیده گشود، اما روش مذکور با آن‌چه که اکنون به‌عنوان ABC می‌شناسیم، دقیقاً معادل نیست. روش‌های ABC برای تقریب تابع توزیع پسین، به جای تابع درست‌نمایی، به کار گرفته می‌شوند. در تحقیقات علوم ژنتیک، اولین بار تاواری و همکاران (۱۹۹۷) الگوریتم اصلی ABC را که حالتی

^۳Smoothing

^۱Intractable

^۲Grid

خاص از روش رد و پذیرش است، معرفی و مورد استفاده قرار دادند. آن‌ها به استنباط در مورد شجره‌نامه افراد با استفاده از داده‌های توالی DNA پرداختند. چنین استنباطی برای هر مدل آماری، رام‌نشدنی است (کورنیوت و همکاران، ۲۰۰۸). مدل‌بندی تغییرات در کروموزوم Y انسان با استفاده از روش ABC توسط پرچارد و همکاران (۱۹۹۹) انجام شد. در نهایت، عبارت محاسبات بیزی تقریبی توسط بیومونت و همکاران (۲۰۰۲) پیشنهاد شد. سایر تعمیم‌های روش ABC و بحث در مورد مناسب بودن این رهیافت، به‌ویژه برای مسایل ژنتیکی به موضوعی بسیار پرتعداد و داغ تبدیل شده است. از آن زمان، ABC کاربردهای دیگری خارج از مطالعات ژنتیکی مانند سیستم‌های زیست‌شناسی^۴، همه‌گیرشناسی^۵، و جغرافیای گیاهی^۶، یافته است. در سال‌های اخیر تعمیم‌های متفاوتی برای روش ABC پیشنهاد و معرفی شده‌اند، که می‌توان به پیتر (۲۰۰۹)، فرنهد و پرانگل (۲۰۱۰)، بیومونت و همکاران (۲۰۱۰)، مارین و همکاران (۲۰۱۲)، مارتین و همکاران (۲۰۱۴) و سیسوسکی (۲۰۱۴) اشاره کرد. در ادامه، به تفصیل به روش‌ها و تعمیم‌هایی که توسط مارین و همکاران (۲۰۱۲) ارائه شده‌اند، خواهیم پرداخت.

۱.۱.۲ انگیزه پیدایش روش‌های ABC

در مدل‌بندی سیستم‌های طبیعی، غالباً فرآیندهای تصادفی پیچیده به فرآیندهای ساده ترجیح داده می‌شوند. علت آن‌هم توانایی در نظر گرفتن و به الگو در آوردن سهم بیشتری از تغییرات منظم پدیده‌های مهم و برجسته یک سیستم است. با این حال، چنین فرآیندهایی می‌توانند به توابع درستنمایی پیچیده‌ای منجر شوند. این چالشی است که بر سر راه روش‌های استنباط آماری مبتنی بر درستنمایی و پیرو آن استنباط بیزی قرار دارد. برای تشریح موضوع، باید متذکر شویم که مهم‌ترین هدف در استنباط بیزی، دانستن توزیع پسین پارامترها به شرط داده‌های مشاهده‌شده است. تابع چگالی این توزیع پسین متناسب است با حاصل ضرب تابع درستنمایی و تابع چگالی پیشین، که به دلیل پیچیده بودن محاسبه تابع درستنمایی محاسبه آن نیز سخت و پیچیده خواهد بود. بنابراین استنباط بیزی که بر پایه توزیع پسین ساخته می‌شود، دچار مشکل و چالش می‌شود. در نتیجه، تحلیل و استنباط بیزی مدلی که شامل تابع درستنمایی رام‌نشدنی است، توجه و نیاز به داشتن راه‌حلی برای گذر از این مشکل را می‌طلبد.

۲.۲ ABC چیست؟

الگوریتم محاسبات بیزی تقریبی، در مواردی که محاسبه تابع درستنمایی سخت و حتی غیرممکن است، مجموعه‌ای از روش‌هایی را ارائه می‌دهد که استنباط بیزی را بدون نیاز به محاسبه صریح تابع درستنمایی مدل، بر اساس یک قاعده مبتنی بر شبیه‌سازی، انجام می‌دهد.

همان‌طور که اشاره کردیم، در غالب موارد، صورت بسته توزیع پسین در دسترس نیست که دلیل آن بسته نبودن صورت تابع درستنمایی مدل و به تبع آن ثابت نرمال‌ساز است. چون در این موارد محاسبه

^۶Phylogeography

^۴Biology

^۵Epidemiology

درست‌نمایی غیر ممکن است یا به انتگرال‌های با بعد بالا منجر می‌شود، به بیان جزئیاتی در مورد این نوع توابع درست‌نمایی می‌پردازیم.

توزیع‌های رام‌نشده، توزیع‌هایی هستند که دارای تابع درست‌نمایی $\pi(y|\theta) = \frac{f(y|\theta)}{C(\theta)}$ می‌باشند، به طوری که رام نبودن خود را از ثابت نرمال‌ساز $C(\theta)$ به ارث می‌برند. برای تقریب این توابع درست‌نمایی از تکنیک‌های مونت کارلوی استاندارد استفاده می‌شود. موری و همکاران (۲۰۰۶)، الگوریتم‌هایی که در این حالت به کار می‌روند را توسعه دادند. الگوریتم ABC، تعمیمی از تکنیک‌های استاندارد مونت کارلویی را برای استفاده در توزیع‌های با توابع درست‌نمایی پیچیده ارائه می‌دهد.

۱.۲.۲ سطوح رام‌نشده در تابع درست‌نمایی

عبارت رام‌نشده در تابع درست‌نمایی می‌تواند در سطوح متفاوتی رخ دهد. دسته‌بندی این سطوح توسط موری و همکاران (۲۰۰۶) ارائه شد و ویلکینسون (۲۰۰۸) با تعمیم این دسته‌بندی به شرح رده‌ای از روش‌های ABC پرداخت.

(الف) یک سطح رام‌نشده تابع درست‌نمایی ناشی از ثابت نرمال‌ساز C است که مستقل از پارامترها می‌باشد. در این حالت درست‌نمایی به صورت $\pi(y|\theta) = \frac{f(y|\theta)}{C}$ است، که $f(y|\theta)$ قسمت معلوم تابع (چگالی) احتمال به ازای مشاهدات y است. در این حالت، مشکل محاسبه ثابت C است که می‌تواند پیچیده و دشوار باشد. البته این سطح پیچیدگی، در روش‌های تقریب توزیع پسین بر اساس نمونه‌گیری، مانند روش‌های نمونه‌گیری MCMC، خللی وارد نمی‌کند. زیرا این روش‌ها تنها نیازمند دانستن $f(y|\theta)$ (بدون ثابت نرمال‌ساز) هستند.

(ب) سطح دیگری از پیچیدگی تابع درست‌نمایی در محاسبه ثابت نرمال‌ساز $C(\theta)$ است که به پارامتر θ وابسته است. در این حالت تابع درست‌نمایی به شکل $\pi(y|\theta) = \frac{f(y|\theta)}{C(\theta)}$ تعریف می‌شود. برای محاسبه درست‌نمایی با پارامتر معین θ ، نیاز هست تا تقریبی از $C(\theta)$ به دست آوریم. معمولاً، ثابت $C(\theta)$ به مدل وابسته است و شامل انتگرال با بعد بالاست، که محاسبه آن را بسیار مشکل می‌سازد. برای محاسبه چنین انتگرال‌هایی می‌توان از روش‌های عددی، مانند روش‌های تربیع‌بندی گاوس هرمیت سازوار^۷ (AGHQ) استفاده کرد، اما روش‌های انتگرال‌گیری عددی به عنوان یک روش استنباط آماری کلیت ندارند. برای مثال، یک تغییر کوچک در ساختار مدل می‌تواند منتج به یک ثابت جدید گردد که با روش عددی طراحی‌شده برای محاسبه ثابت مدل قبل، قابل محاسبه نباشد. همچنین روش‌های عددی توانایی خود را با افزایش بعد از دست می‌دهند و با خطای تقریب بزرگی همراه می‌شوند.

برای اجرای روش‌های نمونه‌گیری مونت کارلویی مانند MCMC، کافی است تابع درست‌نمایی بدون ضریب نرمال‌ساز معلوم باشد. این یعنی ضریب نرمال‌ساز مستقل از پارامترها است (بند الف). بنابراین نمونه‌گیری از توزیع پسین پارامترها زمانی که ثابت نرمال‌ساز، $C(\theta)$ ، تابعی از خود پارامترها است، امکان‌پذیر نیست و بدون اصلاح و تعمیم روش‌های مذکور در این موارد، نمی‌توان

^۷Adaptive Gauss Hermit Quadrature

از آن‌ها استفاده کرد. موری و همکاران (۲۰۰۶) یک روش MCMC مناسب را برای این نوع از سطح پیچیدگی ارایه کردند. در این موارد، روش‌های ABC می‌توانند، به آسانی، نمونه‌گیری از توزیع پسین را انجام دهند؛ با این حال، هدف اصلی روش‌های ABC بسیار باارزش‌تر است. به عبارتی این روش‌ها مشکل رام نبودن تابع درستنمایی در سطوح اصلی را هدف قرار می‌دهد که در قسمت بعدی تشریح شده است.

(ج) در واقع، سطح مهمی از پیچیدگی تابع درستنمایی منجر به تحقق ایده روش ABC شد که در مقاله دیگ و گریتون (۱۹۸۴) به آن پرداخته شده است. این نویسندگان استنباط آماری را در حالت بسامدی و زمانی که فقط مکانیسم تولید داده‌ها معلوم باشد، بررسی کردند. به عبارت دیگر، در این حالت هیچ بخشی از تابع درستنمایی $\pi(y|\theta)$ به روشنی معلوم نیست اما با مفروض بودن پارامتر θ ، مجموعه داده‌ها می‌توانند طبق مدل $\pi(x|\theta)$ تولید شوند. در این صورت، همان‌طور که در انتهای فصل اول اشاره کردیم، روش‌های ABC، استنباط را بر اساس تقریبی از توزیع پسین واقعی پارامترها انجام می‌دهند.

بنابراین برای داشتن صورت تابع چگالی توزیع پسین یا محاسبه تابع درستنمایی، نیاز هست تا توزیع پسین تقریب زده شود. برای این منظور، روشی به کار گرفته می‌شود که ضمن تقریب پسین، دقت بالایی در برآورد نیز به همراه دارد. در واقع، روش ABC با استفاده از شبیه‌سازی، به روش‌های مختلف نمونه‌گیری از پسین می‌پردازد که ضمن کاهش هزینه، با دقت بالایی در برآورد پارامتر همراه است.

۲.۲.۲ کارکرد روش ABC

روش ABC متکی به شبیه‌سازی ارزان از داده‌ها تحت مدل تصادفی است که پارامتر آن مورد بررسی است. با توجه به توزیع پیشین، روش ABC از تعداد زیادی تحقق، $\{(x_1, \theta_1), (x_2, \theta_2), \dots, (x_T, \theta_T)\}$ استفاده می‌کند، که از توزیع توأم داده‌ها و پارامترها تولید شده‌اند، تا تابع درستنمایی را بدون محاسبه آن تقریب بزند.

نحوه عملکرد ABC به این صورت است که با شبیه‌سازی تحقق‌هایی از توزیع توأم داده‌ها و پارامترها، پارامتر θ_i متناظر با مشاهده تولیدشده کمکی $x_i \in D \subset \mathbb{R}^n$ در نظر گرفته می‌شود، که در آن x_i مشابه داده مشاهده‌شده y با درستنمایی متناظر آن است و بنابراین با توزیع پسین متناظر است. برای محاسبه غیرمستقیم تابع درستنمایی $\pi(y|\theta)$ تحت مقادیر تولیدشده θ_i ، حالتی را بررسی می‌کنیم که به مقایسه داده شبیه‌سازی‌شده x_i و داده واقعی y می‌پردازد. اگر به جای داده‌های «همانند» با داده‌های مشاهده‌شده، از داده‌هایی که دقیقاً با داده مشاهده‌شده برابر هستند، استفاده کنیم، روش استنباطی دقیقی رخ می‌دهد (به استثنای خطای مونت کارلو که به دلیل نمونه‌گیری به وجود می‌آید). در این صورت دستور الگوریتم ABC به صورت جدول ۱.۲ است.

به سادگی قابل درک است که این الگوریتم بسیار ناکاراست و برای تولید تحققی مناسب از توزیع پسین، باید تعداد تکرار T بسیار بزرگ باشد. زیرا احتمال برابری دقیق دو بردار مشاهدات برای متغیرهای تصادفی پیوسته، از دیدگاه نظری، برابر صفر است و برای متغیرهای تصادفی گسسته، با

جدول ۱.۲: نمونه‌گیری از توزیع پسین واقعی با الگوریتم ABC

الگوریتم دقیق نمونه‌گیری ABC
برای تکرار $i = 1, \dots, T$
۱. مقادیر (x_i, θ_i) را از توزیع توام مشاهدات و پارامترها تولید کن.
۲. اگر $x_i = y$ آن‌گاه θ_i را به عنوان تحقق‌ی از توزیع پسین بپذیر.

افزایش بعد بردار، به صفر میل می‌کند. بنابراین برای تقریب کاراتر تابع درستنمایی و در نتیجه تولید نمونه از توزیع پسین در روش ABC باید به شکل دیگری عمل کرد. راه‌حل از ارزیابی همانندی بین داده‌های شبیه‌سازی‌شده با داده‌های واقعی ناشی می‌شود. این تشابه باید، به عنوان مثال، به وسیله یک فاصله اقلیدسی بین داده‌های شبیه‌سازی‌شده و داده‌های واقعی اندازه‌گیری گردد. در این حالت نمونه‌های پذیرفته‌شده، تقریبی از توزیع پسین واقعی را در اختیار ما قرار خواهند داد. جدول ۲.۲ صورت کلی الگوریتم را در این حالت نمایش می‌دهد.

جدول ۲.۲: نمونه‌گیری از توزیع پسین با الگوریتم ABC

الگوریتم نمونه‌گیری ABC بر اساس نزدیکی مقادیر تولیدشده به واقعی
برای تکرار $i = 1, \dots, T$
۱. مقادیر (x_i, θ_i) را از توزیع توام مشاهدات و پارامترها تولید کن.
۲. اگر $\rho(x_i, y) < \epsilon$ آن‌گاه θ_i را به عنوان تحقق‌ی از توزیع پسین بپذیر.

این نسخه الگوریتم ABC نیز مشکل مشابه حالت قبلی را دارد. یعنی زمانی که بعد داده‌ها، n ، بالاست، به ندرت همه مولفه‌های بردار x به مولفه‌های بردار y نزدیک می‌شوند، به طوری که شرط $\rho(x, y) < \epsilon$ برقرار شود. بنابراین برای مشاهدات با حجم بالا، که معمولاً هم برقرار است، الگوریتم ناکاراست. این نتیجه باعث شد تا محققین این حوزه جدید، راه‌حل را در کاهش بعد مشاهدات با استفاده از آماره‌های خلاصه‌شده^۸ بیابند. به عبارتی، مقایسه تشابه مشاهدات شبیه‌سازی‌شده با مشاهدات واقعی بر حسب آماره‌های خلاصه‌شده مشاهدات، به جای خود داده‌های اصلی، صورت می‌پذیرد. این ساده‌سازی و کاهش بعد داده‌ها توسط آماره‌های خلاصه‌شده، معمولاً با از دست رفتن بخشی از اطلاعات همراه است. حالت استثناء زمانی است که آماره خلاصه‌شده برای پارامترهای مورد نظر، بسنده باشد. با این توصیف، اگر آماره خلاصه‌شده را با $S(\cdot)$ نشان دهیم، الگوریتم ABC به صورت جدول ۳.۲ خواهد بود.

تا این جا مشخص می‌شود که برای یک مدل معین، در کنار توزیع پیشین، سه مولفه تاثیرگذار یک الگوریتم ABC معمول، همان‌هایی هستند که در فصل اول تشریح کردیم: تابع نرم (فاصله) $\rho(\cdot)$ ، آماره

^۸Summary statistics

جدول ۳.۲: نمونه‌گیری از توزیع پسین با الگوریتم ABC و بر حسب آماره خلاصه‌شده

الگوریتم نمونه‌گیری ABC بر اساس آماره خلاصه‌شده
برای تکرار $i = 1, \dots, T$
۱. مقادیر (x_i, θ_i) را از توزیع توام مشاهدات و پارامترها تولید کن.
۲. اگر $\rho(S(x_i), S(y)) < \epsilon$ آن‌گاه θ_i را به عنوان تحقق‌ی از توزیع پسین بپذیر.

خلاصه‌شده $S(\cdot)$ و سطح تحمل ϵ .

نمونه‌گیری از توزیع توام داده‌ها و پارامترها را با استفاده از قانون ضرب احتمال و تعریف احتمال شرطی می‌توان ساده کرد. این توزیع توام را می‌توان به صورت

$$\pi(y, \theta) = \pi(\theta)\pi(y|\theta),$$

نوشت. بنابراین برای شبیه‌سازی می‌توان به صورت سلسله‌مراتبی عمل کرد. به این معنی که ابتدا پارامترها را از توزیع پیشین $\pi(\cdot)$ شبیه‌سازی کنیم و سپس داده‌ها را تحت توزیع شرطی مشاهدات به شرط پارامترها، $\pi(\cdot|\theta)$ ، یا همان تابع درست‌نمایی با پارامتر معلوم، شبیه‌سازی کنیم. معمولاً شبیه‌سازی پارامترها از توزیع پیشین راحت است، زیرا معمولاً توزیع پیشین را کاملاً معلوم فرض می‌کنیم. در مقابل، شبیه‌سازی داده‌ها تحت یک مدل شرطی، به شرط مفروض بودن پارامترها، همیشه آسان نیست؛ زیرا مکانیسم تصادفی تولید داده‌ها می‌تواند پیچیده باشد. برای این قسمت معمولاً از روش‌های نمونه‌گیری مطرح‌شده در فصل اول استفاده می‌شود که منجر به پیدایش صورت‌های مختلفی از الگوریتم‌های ABC شده است و در ادامه به معرفی برخی از آن‌ها می‌پردازیم.

برای داده‌ها و پارامترهای با بعد بالا، غالب داده‌های شبیه‌سازی‌شده غیرمشابه با داده‌های واقعی، رد خواهند شد. این شبیه‌سازی‌ها باعث می‌شود تا نمونه‌گیری از پسین بی‌فایده گردد. بنابراین برای به‌دست آوردن اندازه نمونه مناسب برای تقریب خوب توزیع پسین، T باید بزرگ انتخاب شود که این خود هزینه بالایی در پی خواهد داشت.

۳.۲ الگوریتم‌های مختلف ABC

در این بخش به معرفی سه الگوریتم اصلی نمونه‌گیری مبتنی بر ABC، می‌پردازیم. ابتدا ABC استاندارد یا الگوریتم رد-ABC (RS-ABC) را بیان می‌کنیم و با مثالی عملکرد آن را تشریح می‌کنیم. الگوریتم نمونه‌گیری رد-ABC در فهم و پیاده‌سازی آسان است و به‌طور گسترده استفاده می‌شود. اما در عین پرطرفدار بودن، زمانی که وزن پسین سهم بالایی را از پیشین اختیار کند، می‌تواند ناکارا باشد؛ به عبارتی اکثر نمونه‌های پیشنهادی رد می‌شوند. به عنوان رهیافتی جانشین، دو الگوریتم مونت کارلو (نمونه‌گیری MCMC و نقاط مهم) که در جستجوی فضای پارامتر، برای استنباط ABC، موثرتر هستند را در این بخش معرفی می‌کنیم. این دو الگوریتم نسبت به الگوریتم ABC استاندارد دارای مزایایی هستند و در

مقابل پیچیدگی اجرای بیشتری دارند.

۱.۳.۲ نمونه‌گیری رد-ABC

فرض کنید y داده مشاهده‌شده تحت مدل $\pi(y|\theta)$ باشد که θ برداری از پارامترها است. برای سادگی فرض کنید θ یک پارامتر عددی است و در مدل پارامتر مزاحم وجود ندارد. البته در بیشتر موقعیت‌های کاربردی مدل‌های تصادفی دارای پارامتری با بعد بالا هستند و با پارامتر مزاحم نیز همراه می‌باشند. برای رفع این گرفتاری معمولاً از مدل نسبت به پارامترهای مزاحم انتگرال‌گیری می‌کنند و در مواردی که محاسبه انتگرال پیچیده و سخت است از روش‌های انتگرال‌گیری مونت کارلویی یا نمونه‌گیری از توزیع توام پسین همه پارامترها استفاده می‌کنند. اگر تابع چگالی توزیع پسین θ ، $\pi(\theta|y)$ ، به صورت (۱.۱) باشد، آنگاه آن را به وسیله روش محاسبات استاندارد همچون نمونه‌گیری رد-ABC می‌توان تقریب زد.

روش‌های ABC از توزیع پسین بدون محاسبه درست‌نمایی، نمونه‌گیری می‌کند. تاواری و همکاران (۱۹۹۷) نمونه‌گیری را با تولید تحقق از $\pi(y|\theta)$ انجام دادند و پارامتر θ را با تولید تحقق‌های مستقل $\{(x_1, \theta_1), (x_2, \theta_2), \dots, (x_T, \theta_T)\}$ از توزیع توام $\pi(x, \theta)$ شبیه‌سازی کردند. در این روش، توزیع هدف را می‌توان به صورت توزیع حاشیه‌ای مدل $\pi(\theta, x|y)$ نسبت به x ، در نظر گرفت. فرض کنید $x \in D$ داده‌های شبیه‌سازی‌شده باشند که در آن D همان تکیه‌گاه بردار مشاهدات y است. می‌توان نوشت

$$\begin{aligned} \int \pi(\theta, x|y) dx &\propto \int \pi(y|\theta, x) \pi(\theta, x) dx \\ &= \int \mathbb{I}_{(x=y)} \pi(x|\theta) \pi(\theta) dx \end{aligned} \quad (1.2)$$

که در آن $\mathbb{I}_{(x=y)}$ تابع نشانگری است که به ازای $x = y$ مقدار یک و در سایر موارد مقدار صفر را اختیار می‌کند. روش تاواری و همکاران (۱۹۹۷)، نشان داد که در کنار انتخاب توزیع پیشین، که در استنباط بیزی مورد نیاز است، مشاهدات شبیه‌سازی‌شده از توزیع توام داده‌ها و پارامترها، برای نمونه‌گیری از توزیع پسین الزامیست.

همان‌طور که پیشتر متذکر شدیم، متأسفانه، این الگوریتم تولید نمونه از توزیع پسین را برای مدل‌های پیچیده، سخت‌تر می‌کند. دلیل آن هم ناشی از دو منبع است؛ اول این‌که برای داده‌های با بعد بالا، بررسی شرط $x = y$ برای همه داده‌های شبیه‌سازی‌شده واقعا ممکن نیست. اشاره کردیم که جایگزینی داده‌ها با آماره‌های خلاصه‌شده، یک روش رایج برای حل این مشکل است. بنابراین داده‌های مشاهده‌شده y و شبیه‌سازی‌شده x ، به ترتیب با $S(y) = s_y$ و $S(x) = s_x$ جایگزین می‌شوند. اگرچه به واسطه این خلاصه‌سازی قسمتی از اطلاعات موجود از دست می‌رود و سطحی از خطا در استنباط مبتنی بر ABC وارد می‌شود، اما در اکثر موارد این نقص اطلاعات زیاد نیست و حتی برای رده بزرگی از مدل‌های آماری (یعنی خانواده توزیع‌های نمایی) می‌توان امیدوار بود آماره‌های خلاصه‌شده بسنده باشند. توصیف رابطه بین $\pi(\theta|x)$ و $\pi(\theta|s)$ ، یک پژوهش مورد بررسی در این زمینه است (پرانگل، ۲۰۱۱؛ ویلکینسون، ۲۰۰۸).

دوم، حتی موقعی که توزیع پسین $\pi(\theta|s_y)$ هدف باشد، تساوی $s_y = s_x$ ممکن است اتفاق نیافتد. برای مثال، زمانی که داده‌ها پیوسته هستند، احتمال این برابری مساوی صفر است. همان‌طور که قبلاً گفتیم، برای رفع این مشکل به جای شرط پذیرش برابری اکید $s_y = s_x$ ، یک فاصله تحمل $\rho_\epsilon(s_y, s_x)$ تعریف می‌کنند که در آن ϵ همان سطح تحمل یا پهنای نوار است. این رهیافت نیز سطح دیگری از خطا را وارد مساله استنباط می‌کند. دو سطح تقریب مطرح شده، توزیع پسینی از پارامترها را به وسیله ABC می‌دهد که به صورت زیر تعریف می‌شود:

$$\pi_\epsilon(\theta|y) \propto \int \rho_\epsilon(s_y, s_x) \pi(s_x|\theta) \pi(\theta) dx. \quad (2.2)$$

با توجه به عبارت (۲.۲)، پریچارد و همکاران (۱۹۹۹) الگوریتم رد-ABC را به صورت جدول ۴.۲ ارائه دادند، که یک نسخه تقریبی از روش تاواری و همکاران است.

جدول ۴.۲: نمونه‌گیری RS-ABC

الگوریتم نمونه‌گیری RS-ABC
برای $i = 1, \dots, T$ تکرار کن
۱- θ' را از توزیع پیشین $\pi(\cdot)$ تولید کن.
۲- x را از درست‌نمایی $\pi(\cdot \theta')$ تولید کن.
۳- اگر $\rho(S(x), S(y)) \leq \epsilon$ ، آن‌گاه قرار بده $\theta_i = \theta'$.
در غیر این صورت θ' را رد کن و به مرحله ۱ بازگرد.
خروجی: مجموعه‌ای از پارامترهای پذیرفته شده

در الگوریتم نمونه‌گیری ۴.۲

- $S(\cdot)$: آماره خلاصه شده تعریف شده روی D است که غالباً بسنده نیست.

- $\rho(\cdot)$: تابع فاصله روی $S(D)$.

- ϵ : سطح تحملی که همیشه مثبت است.

در واقع، الگوریتم ۴.۲ از پسینی به شکل زیر که بر اساس متغیر کمکی x شکل می‌گیرد، نمونه‌گیری می‌کند:

$$\pi_\epsilon(\theta, x|y) = \frac{\pi(\theta) \pi(x|\theta) \mathbb{I}_{A_{\epsilon,y}}(x)}{\int_{A_{\epsilon,y}} \pi(\theta) \pi(x|\theta) dx d\theta}, \quad (3.2)$$

که در آن $\mathbb{I}_B(\cdot)$ تابع نشانگر مجموعه B است و $\mathbb{I}_{A_{\epsilon,y}} = \{x \in D | \rho[s_x, s_y] \leq \epsilon\}$. ایده اصلی ABC این است که: با استفاده از آماره (به اندازه کافی) خلاصه شده $S(\cdot)$ ، همراه با تحمل (به اندازه کافی) کوچک

ϵ ، می‌بایست بتوان تقریبی (به اندازه کافی) مناسب از توزیع پسین ایجاد کرد. یعنی

$$\pi_{\epsilon}(\theta|y) = \int \pi_{\epsilon}(\theta, x|y)dx \approx \pi(\theta|y). \quad (۴.۲)$$

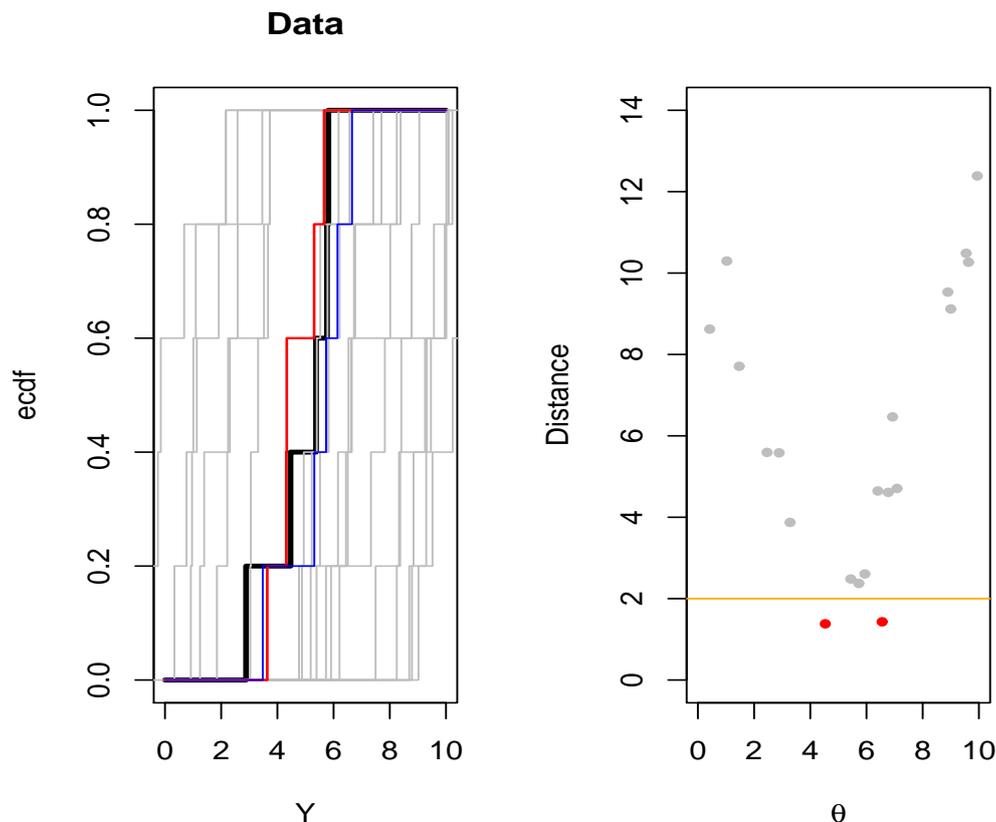
پارامترهای پذیرفته‌شده توسط الگوریتم ۴.۲ دارای تابع چگالی پسین متناسب با (۲.۲) هستند و اگر $\epsilon \rightarrow 0$ ، آن‌گاه به پسین $\pi(\theta|y)$ همگرا می‌شود. با این حال، در اکثر مدل‌های مورد بررسی، داده‌ها پیوسته هستند و بنابراین احتمال پذیرش، وقتی $\epsilon \rightarrow 0$ ، صفر است. در بیشتر مواقع $\epsilon > 0$ در نظر گرفته می‌شود. در حال حاضر توزیع هدف، تقریبی از توزیع پسین واقعی را می‌دهد که این تقریب وابسته به انتخاب $\rho(\cdot)$ و به‌ویژه ϵ است. برای روشن شدن مطلب، مثالی را بررسی می‌کنیم.

مثال ۱.۳.۲. مدلی را که دارای پنج مشاهده مستقل از توزیع $N(\theta, 1)$ است، بررسی می‌کنیم. در این مدل فرض می‌کنیم θ دارای توزیع پیشین یکنواخت در فاصله $(0, 1)$ است. توجه کنید که این مدل و برازش آن ساده است، یعنی تابع درستنمایی و به تبع آن ثابت نرمال‌سازش به‌سادگی قابل محاسبه هستند. در نتیجه نیازی به استفاده از ABC نیست. اما برای تشریح عملکرد روش ABC، این مثال ساده را مطرح کردیم. برای نمایش مکانیسم پذیرش، شکل ۱.۲ جزئیات بیشتری از الگوریتم را نشان می‌دهد. نتایج بر اساس $N = 20$ مشاهده واقعی و شبیه‌سازی شده به‌دست آمده‌اند. تابع فاصله برابر $\rho(x, y) = \|x - y\|_2$ در نظر گرفته شد. در نمودار سمت چپ شکل ۱.۲ داده‌های مشاهده‌شده با خط مشکی و داده‌های شبیه‌سازی شده با رنگ خاکستری نمایش داده شده‌اند. دو تحقق پذیرفته‌شده رنگی هستند. نمودار سمت راست شکل ۱.۲ مقادیر پارامترهای شبیه‌سازی شده را نشان می‌دهد که دو پارامتر پذیرفته‌شده با رنگ قرمز و باقی پارامترها که رد شده‌اند، خاکستری رنگ هستند. سطح تحمل ϵ ، خط نارنجی است که در نمودار سمت راست مشخص است.

شکل ۲.۲، برآورد تابع چگالی پسین (که متناسب با هسته نرمال است) را با توجه به مقادیر مختلف ϵ نمایش می‌دهد. در این شکل $N = 1000$ در نظر گرفته شد. نمودار سمت چپ شکل ۲.۲ برآورد چگالی پسین حاصل از خروجی با خط پرنرنگ و پسین واقعی به صورت نقطه چین است. نمودار سمت راست شکل ۲.۲، پارامترهای شبیه‌سازی شده را نشان می‌دهد. مقادیر پذیرفته‌شده قرمز هستند و مقادیر رد که بالاتر از خط ϵ قرار گرفته‌اند، خاکستری رنگ هستند. همچنین سطرها از بالا به پایین، به ترتیب متناظر با مقادیر $0.1, 0.05, 0.1, 0.25$ برای ϵ هستند. بررسی نمودارهای شکل ۲.۲ نشان می‌دهد که بهترین تقریب برای سطر دوم و سوم با $0.2 \approx \epsilon$ حاصل می‌شود. برای کوچکترین ϵ ، به علت حجم کوچک نمونه پذیرفته‌شده، خطای برآورد زیاد است و این همان ناکارایی مطرح‌شده در بخش‌های قبلی است. برای بزرگترین مقدار ϵ ، حجم بیشتری از پارامترهای شبیه‌سازی شده نسبت به بقیه حالت‌ها پذیرفته می‌شوند. این پیامد واریانس پارامترهای پذیرفته‌شده را افزایش می‌دهد. در عمل یک داد و ستد^۹ بین میزان خطا و ϵ وجود دارد، که باید بر اساس چارچوبی تعیین گردند.

کارایی الگوریتم را می‌توان با استفاده از آماره‌های خلاصه‌شده (بسند) افزایش داد. شکل ۳.۲ نتیجه الگوریتم RS-ABC بر اساس آماره بسند \bar{y} است. به‌طور کلی با استفاده از آماره بسند (در صورت وجود) و الگوریتم نمونه‌گیری ۴.۲، توزیع حدی وقتی ϵ کوچک است، تقریباً همان توزیع پسین واقعی

^۹Trade-off

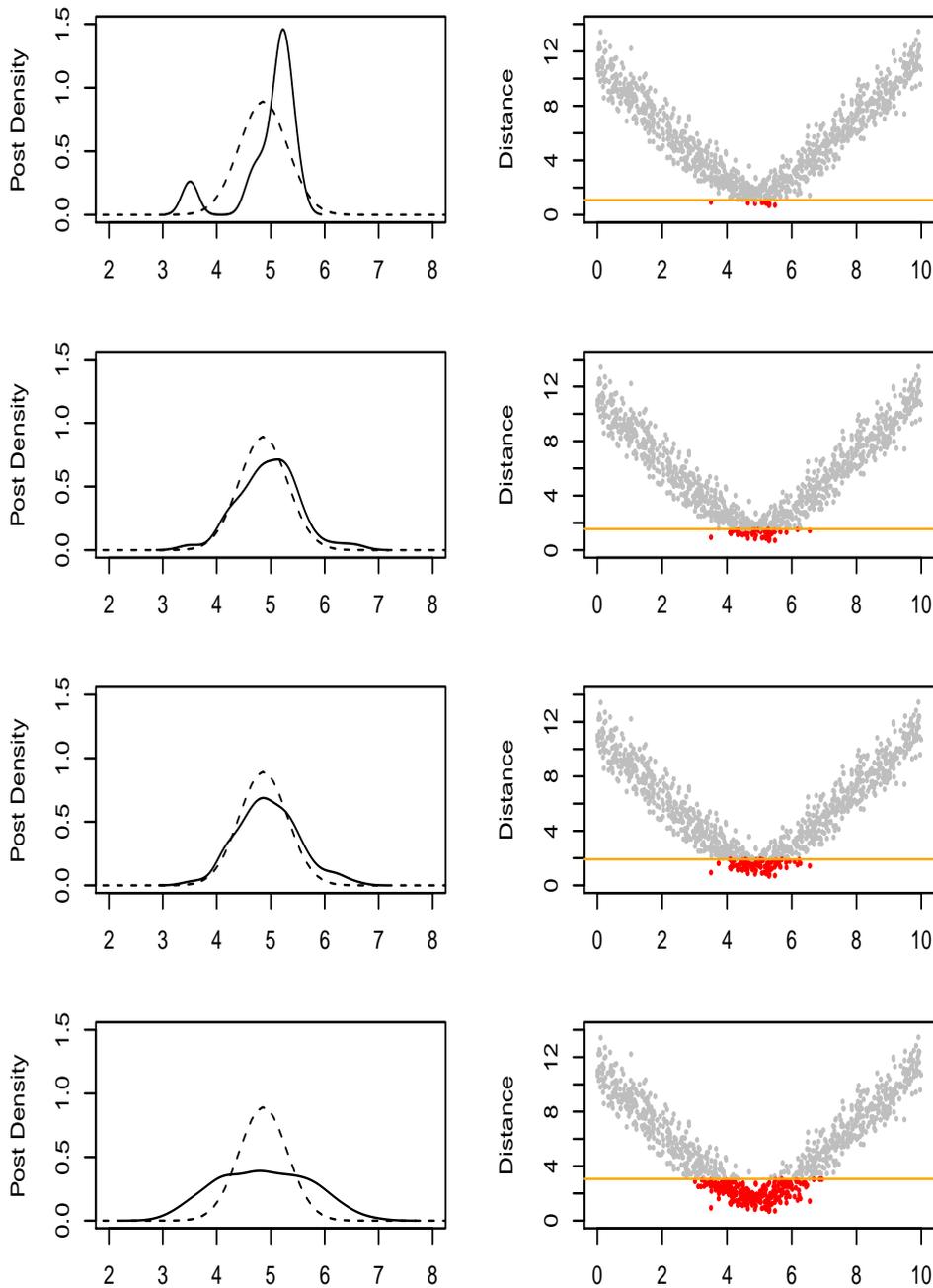


شکل ۱۰.۲: خروجی برای 20 تکرار الگوریتم ۴.۲ بر اساس مدل نرمال است. نمودار سمت چپ توزیع تجمعی تجربی داده واقعی و شبیه‌سازی شده و سمت راست مقادیر پذیرفته را نشان می‌دهد.

است. هر سطر نمودار ۳.۲ از تعداد مشابه داده‌های شبیه‌سازی شده متناظر با سطرهای شکل ۲.۲ استفاده می‌کند. اگر این تعداد کاهش یابد منتج به دقت بالا می‌شود. کاهش دقت الگوریتم ABC متناظر با افزایش بعد در آماره خلاصه شده است.

پیغامی که از شکل ۳.۲ فرستاده می‌شود، این است که استفاده از آماره خلاصه شده بسنده با بعد کم می‌تواند کارایی الگوریتم را افزایش دهد، که در نمودارهای این شکل به خوبی مشهود است. اما در موقعیت‌های واقعی، معمولاً با داده‌های با بعد بالا مواجه هستیم و هیچ آماره بسنده‌ای برای پارامترهای مورد علاقه نمی‌توان یافت. در این صورت باید از آماره‌های خلاصه شده (غیربسند) با بعد کم استفاده کرد. البته با این کار مقداری از اطلاعات از دست می‌رود، که باعث کاهش کارایی الگوریتم می‌شود. با این حال، روش‌هایی موجودند که بهترین آماره خلاصه شده را انتخاب می‌کنند.

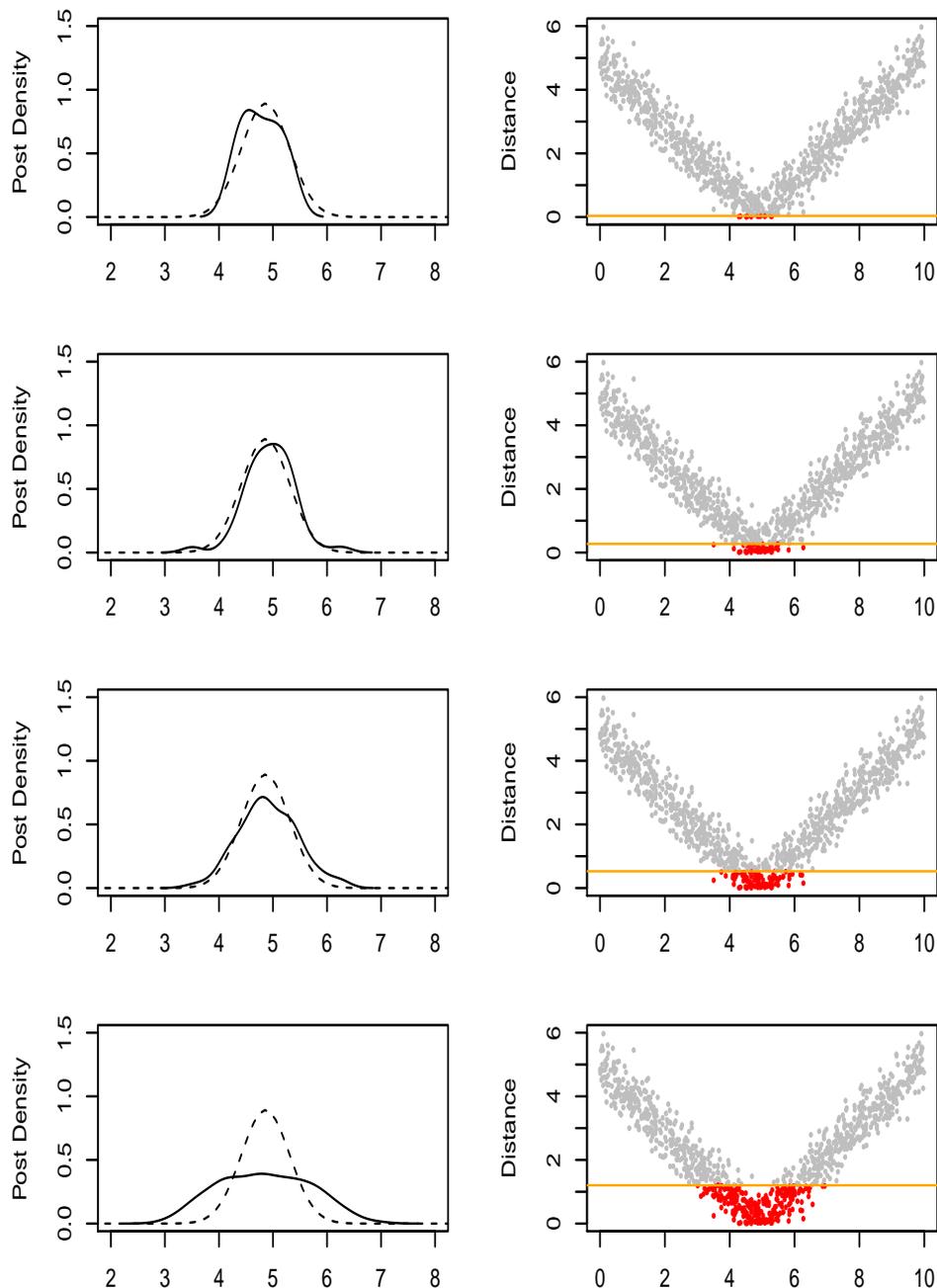
وقتی از سطوح متفاوتی برای تقریب توزیع پسین استفاده می‌شود، خطاهای تقریبی در ABC به وجود می‌آید: یک خطا ناشی از نمونه‌گیری مونت کارلویی است؛ خطای دوم با از دست دادن اطلاعات به دلیل استفاده از $S(\cdot)$ (که معمولاً بسنده نیست) به وجود می‌آید؛ و خطایی که در برآورد تابع چگالی هدف به علت استفاده از $\epsilon > 0$ به وجود می‌آید.



شکل ۳.۲: خروجی الگوریتم ۱.۲ بر اساس ۱۰۰۰ تکرار از توزیع نرمال است. سطرها از بالا به پایین، به ترتیب متناظر با مقادیر ۰/۱، ۰/۰۵، ۰/۰۱ و ۰/۰۲۵ برای ϵ هستند. در تمام نمودارها محور افقی θ ها هستند.

تابع چگالی توزیع پسین ABC را با توجه به تعریف (۱.۱) و استفاده از آماره‌های خلاصه‌شده، می‌توان به صورت زیر نوشت:

$$\pi_{ABC}(\theta|s_y) \propto \pi(s_y|\theta)\pi(\theta), \quad (۵.۲)$$



شکل ۳.۲: نمودارهای مشابه شکل ۲.۲ که بر اساس آماره بسنده $S(y) = \bar{y}$ نتیجه شده‌اند

که در آن

$$\pi(s_y|\theta) = \int \pi(y|\theta) \mathcal{K}\left(\frac{s_x - s_y}{\epsilon}\right) dx. \quad (۶.۲)$$

عبارت (۵.۲)، تابع چگالی پسین ABC است و عبارت (۶.۲) به‌عنوان تقریب ABC برای تابع درست‌نمایی محسوب می‌شود.

قبل از پایان این زیربخش، نکته مهمی که باید به آن توجه کرد آن است که در الگوریتم‌های ABC،

در عمل استفاده از توزیع پیشین ناآگاهی بخش^{۱۰}، بسیار ناکارا است، زیرا این توزیع‌های پیشین ممکن است مقادیری از فضای پارامتر را پیشنهاد کنند که در نواحی کم‌چگال توزیع پسین قرار گیرند. بنابراین در این حالت، مقادیر پیشنهادی از توزیع پسین یا به‌طور یکنواخت تولید می‌شوند یا از قسمت‌هایی که تراکم چگالی پسین واقعی در آن‌ها کم است، انتخاب می‌شوند.

۲.۳.۲ نمونه‌گیری نقاط مهم ABC

جدول ۵.۲ یک الگوریتم نمونه‌گیری نقاط مهم ABC (IS-ABC) را نشان می‌دهد. خروجی الگوریتم یک نمونه به حجم T از پارامترهای مدل و وزن‌های متناظر با آن‌هاست؛ یعنی $\{(\theta_i, w_i)\}_{1 \leq i \leq T}$. با این خروجی هر تابعی مانند $h(\theta)$ را می‌توان به صورت

$$\frac{\sum_{i=1}^T h(\theta_i)w_i}{\sum_{i=1}^T w_i}, \quad (۷.۲)$$

برآورد کرد. لیو (۱۹۹۶) و پرانگل (۲۰۱۱) نشان دادند که برای T بزرگ، برآوردگر (۷.۲) در احتمال به

$$\begin{aligned} \frac{E(h(\theta)w)}{E(w)} &= \frac{E(h(\theta)E(w|\theta))}{E(w|\theta)} \\ &= \frac{\int h(\theta)E(w|\theta)g(\theta)d\theta}{\int E(w|\theta)g(\theta)d\theta}, \end{aligned} \quad (۸.۲)$$

همگرا می‌گردد. عبارت (۸.۲) امید ریاضی $h(\theta)$ تحت تابع چگالی متناسب با $E(w|\theta)g(\theta)$ است. برای جدول ۵.۲، این چگالی مساوی با تابع چگالی پسین ABC است. در مورد الگوریتم IS-ABC چند نکته زیر قابل طرح هستند:

- تعدادی از وزن‌های الگوریتم جدول ۵.۲ صفر می‌شوند و در عمل مقادیر θ_i متناظر با این وزن‌ها از نمونه حذف خواهند شد.
- یک حالت خاص الگوریتم IS-ABC در جدول ۵.۲، وقتی است که $g(\theta) = \pi(\theta)$. در این حالت نتیجه همان الگوریتم نمونه‌گیری RS-ABC خواهد بود.
- در گام ۳ الگوریتم، از توابع هسته متفاوت می‌توان استفاده کرد. یک انتخاب معمول، که استفاده از آن در الگوریتم نمونه‌گیری RS-ABC رایج است، تابع هسته یکنواخت می‌باشد. چنانچه $\mathcal{K}(\cdot)$ یکنواخت نباشد، می‌توان با تغییر گام ۳ یک الگوریتم IS-ABC دیگر ساخت. در این صورت، تابع وزن به شکل

$$w = \mathcal{K}\left(\frac{s_x - s_y}{\epsilon}\right) \frac{\pi(\theta')}{g(\theta')} \quad (۹.۲)$$

تعریف می‌شود. با استدلالی مشابه می‌توان نشان داد، با این الگوریتم توزیع هدف همان (۱.۱) است.

^{۱۰} Noninformative priors

- با انتخاب مناسب $g(\cdot)$ ، واریانس برآوردگر (۷.۲) می‌تواند کوچک شود.

جدول ۵.۲: نمونه‌گیری IS-ABC

الگوریتم نمونه‌گیری نقاط مهم ABC
برای تکرار $i = 1, \dots, T$
۱. θ' را از توزیع پیشنهادی $g(\cdot)$ تولید کن.
۲. x را از درستیابی $\pi(\cdot \theta')$ تولید کن.
۳. با احتمال $\mathcal{K}((s_x - s_y)/\epsilon)$ ، قرار بده $w_i = \pi(\theta')/g(\theta')$ ، در غیر این صورت قرار بده $w_i = 0$.
خروجی: مجموعه‌ای از مقادیر θ_i به همراه وزن‌های متناظر w_i .

۳.۳.۲ نمونه‌گیری مونت کارلوی زنجیر مارکوفی ABC

الگوریتم مونت کارلوی زنجیر مارکوفی ABC (MCMC-ABC) در جدول ۶.۲ ارائه شده است. ماجروم و همکاران (۲۰۰۳) این الگوریتم را برای حالتی که هسته انتقال یکنواخت است، مطرح و ثابت کردند چگالی هدف تابع چگالی (۵.۲) است. انگیزه معرفی این نسخه الگوریتم، استفاده از توزیع‌های پیشین ناآگاهی‌بخش بود. زیرا، همان‌طور که قبلاً متذکر شدیم، استفاده از این پیشین‌ها در الگوریتم ABC می‌تواند الگوریتم را بسیار ناکارآمد سازد. دلیل آن هم تولید مقادیر پیشنهادی است که در نواحی پسینی کم‌چگال (با احتمال کوچک) قرار می‌گیرند. سیزن و فان (۲۰۱۰)، با استفاده از MCMC، نسخه دیگری برای ABC ارائه دادند و الگوریتم را تعمیم دادند.

در مقایسه با روش RS-ABC در این نسخه ABC، با استفاده از الگوریتم MCMC، به‌جای استفاده مستقیم از توزیع پیشین از هسته مارکوف $k(\cdot|\cdot)$ استفاده می‌کنیم. یعنی با تشکیل یک زنجیر مارکوف که به توزیع مانای خود (توزیع پسین) همگرا باشد، تقریبی برای توزیع پسین، بدون محاسبه تابع درستیابی، به‌دست می‌آوریم. برای این منظور از $k(\cdot|\theta^i)$ تولید نمونه می‌کنیم و با استفاده از قضیه ارگودیک توزیع پسین را به شکل زیر تقریب می‌زنیم:

$$\pi(\theta|y) \approx \frac{\pi(\theta)}{T} \sum_{i=1}^T \pi(y|x_i, \theta), \quad (10.2)$$

که در آن x مجموعه‌ای از داده‌های مستقل و دارای چگالی $\pi(x|\theta)$ است (فصل یک را ببینید). در الگوریتم MCMC-ABC، نمونه‌گیری MCMC، و به‌طور خاص نمونه‌گیری متروپولیس-هستینگر، برای تولید نمونه‌ای از توزیع هدف ABC، $\pi_{ABC}(\theta, x|y)$ ، بدون استفاده از درستیابی رام‌نشده طراحی شده است (ماجروم و همکاران، ۲۰۰۳). در این الگوریتم، توزیع پیشنهادی به صورت

زیر است:

$$k[(\theta', x') | (\theta, x)] = k(\theta' | \theta) \pi(x' | \theta'). \quad (11.2)$$

در واقع، هنگامی که الگوریتم در حالت (θ, x) است، مقدار جدید θ' از توزیع پیشنهادی $k(\theta' | \theta)$ شبیه‌سازی می‌شود. سپس با شرط $\theta = \theta'$ ، مقدار جدید x' از مدل $\pi(\cdot | \theta')$ تولید می‌شود؛ به عبارتی گام دوم الگوریتم ۶.۲ انجام می‌شود. بنابراین زنجیر از (θ, x) به (θ', x') انتقال می‌یابد. با این استدلال به یک زنجیر مارکوف با توزیع مانای $\pi_{ABC}(\theta, x | y)$ دست پیدا می‌کنیم. برای رسیدن به این توزیع مانا، باید تاکید خود را روی شرط تعادل دقیق بگذاریم. یعنی

$$\pi_{ABC}(\theta, x | y) p[(\theta', x') | (\theta, x)] = \pi_{ABC}(\theta', x' | y) p[(\theta, x) | (\theta', x')], \quad (12.2)$$

که در آن احتمال انتقال MCMC (متروپولیس-هستینگز) به صورت زیر تعریف می‌شود:

$$p[(\theta', x') | (\theta, x)] = k[(\theta', x') | (\theta, x)] \alpha[(\theta, x), (\theta', x')], \quad (13.2)$$

که در آن α احتمال پذیرش الگوریتم MCMC-ABC در حالتی است که کمتر از ۱ باشد. ایده ABC ارابه تقریبی برای توزیع پسین (۵.۲) است. دقت این تقریب به انتخاب $S(\cdot)$ ، $\mathcal{K}(\cdot)$ و ϵ بستگی دارد. برای مثال، اگر $S(\cdot)$ تابع همانی باشد و $\epsilon \rightarrow 0$ ، آن‌گاه پسین ABC به پسین واقعی همگرا می‌گردد. این همگرایی برای حالتی که $S(\cdot)$ آماره بسنده پارامتر مورد نظر باشد نیز اتفاق می‌افتد.

جدول ۶.۲: نمونه‌گیری MCMC-ABC

الگوریتم نمونه‌گیری MCMC-ABC
با داشتن تحقق $(\theta^{(0)}, s^{(0)})$ از توزیع هدف $\pi_\epsilon(\theta, x y)$
برای $i = 1, \dots, T$ تکرار کن
۱- θ^{i-1} را از هسته مارکوف $k(\cdot \theta^{(i-1)})$ تولید کن.
۲- x را از درستی‌نمایی $\pi(\cdot \theta')$ تولید کن.
۳- با احتمال $\min \left\{ 1, \frac{\mathcal{K}((s_x - s_y)/\epsilon) \pi(\theta) k(\theta^{(i-1)} \theta)}{\mathcal{K}((s_x^{(i-1)} - s_y)/\epsilon) \pi(\theta^{(i-1)}) k(\theta \theta^{(i-1)})} \right\}$ ، θ را بپذیر و قرار بده $\theta^i = \theta$ و $s_x^i = s_x$.
در غیر این صورت قرار بده $\theta^i = \theta^{(i-1)}$ و $s_x^i = s_x^{(i-1)}$.
خروجی: مجموعه‌ای از مقادیر θ^i

ممکن است این سوال در ذهن خواننده ظاهر شود که نقش $\rho(s_x, s_y) \leq \epsilon$ در الگوریتم ۶.۲ چیست؟

در پاسخ تاکید می‌کنیم که $\rho(s_x, s_y) \leq \epsilon$ همان $\mathcal{K}((s_x - s_y)/\epsilon)$ است.

احتمال استفاده شده در الگوریتم ۶.۲ شامل محاسبه تابع درستنمایی نیست و بنابراین به لحاظ محاسبه برای ABC سودمند است. با استفاده از الگوریتم نمونه‌گیری MCMC-ABC، نمونه‌ای تولید می‌شود که تقریباً دارای توزیع هدف $\pi_\epsilon(\theta, x|y)$ است. احتمال پذیرش الگوریتم ۶.۲، یعنی احتمال حرکت زنجیر از $(\theta^{(i-1)}, x^{(i-1)})$ به (θ', x') ، در مرحله i ام، در چارچوب الگوریتم MCMC (متروپولیس-هستینگز)، برابر $\min\{1, \alpha[(\theta^{(i-1)}, x^{(i-1)}), (\theta', x')]\}$ است. در واقع، با توجه به (۱۳.۱)، (۳.۲)، (۱۱.۲) و شکل توزیع پسین ABC، داریم

$$\begin{aligned} \alpha[(\theta^{(i-1)}, x^{(i-1)}), (\theta', x')] &= \frac{\pi_{ABC}(\theta', x'|y)}{\pi_{ABC}(\theta^{(i-1)}, x^{(i-1)}|y)} \times \frac{k[(\theta^{(i-1)}, x^{(i-1)})|(\theta', x')]}{k[(\theta', x')|(\theta^{(i-1)}, x^{(i-1)})]} \\ &= \frac{\pi_\epsilon(\theta', x'|y)}{\pi_\epsilon(\theta^{(i-1)}, x^{(i-1)}|y)} \times \frac{k(\theta^{(i-1)}|\theta')\pi(x^{(i-1)}|\theta^{(i-1)})}{k(\theta'|\theta^{(i-1)})\pi(x'|\theta')} \\ &= \frac{\pi(\theta')\pi(x'|\theta')\mathbb{I}_{A_{\epsilon,y}}(x')}{\pi(\theta^{(i-1)})\pi(x^{(i-1)}|\theta^{(i-1)})\mathbb{I}_{A_{\epsilon,y}}(x^{(i-1)})} \\ &\quad \times \frac{k(\theta^{(i-1)}|\theta')\pi(x^{(i-1)}|\theta^{(i-1)})}{k(\theta'|\theta^{(i-1)})\pi(x'|\theta')} \\ &= \frac{\pi(\theta')k(\theta^{(i-1)}|\theta')}{\pi(\theta^{(i-1)})k(\theta'|\theta^{(i-1)})} \underbrace{\mathbb{I}_{A_{\epsilon,y}}(x')}_{\frac{\pi_\epsilon(y|x',\theta')}{\pi_\epsilon(y|x,\theta)}}, \end{aligned} \quad (۱۴.۲)$$

که در آن $\mathbb{I}_{A_{\epsilon,y}}$ در رابطه (۳.۲) تعریف شده است. با توجه به عبارت (۱۴.۲) و احتمال پذیرش نتیجه‌شده، نیازی به محاسبه مستقیم تابع درستنمایی رام‌نشده نیست. بنابراین یک محاسبه رام‌شدنی داریم که می‌تواند به آسانی انجام شود. اگر تابع تصمیم را به صورت

$$\pi_\epsilon(y|x, \theta) = \begin{cases} 1 & \text{اگر } \rho(s(x), s(y)) \leq \epsilon \\ 0 & \text{در غیر این صورت} \end{cases} \quad (۱۵.۲)$$

تعریف کنیم، با توجه به (۳.۲) تابع نشانگر $\mathbb{I}_{A_{\epsilon,y}}(x')$ به جای تابع درستنمایی $\pi_\epsilon(y|x', \theta')$ قرار می‌گیرد.

با تعریف احتمال پذیرش به صورت $\min\{1, \alpha[(\theta, x), (\theta', x')]\}$ ، نشان می‌دهیم شرط تعادل دقیق برقرار است. قبلاً اشاره کردیم که برقراری این شرط همگرایی به توزیع مانای هدف ABC را تضمین

می‌کند. با استفاده از (۱۳.۲) و (۱۴.۲) می‌توان نوشت

$$\begin{aligned}
 \pi_{ABC}(\theta, x|y)p[(\theta', x')|(\theta, x)] &= \pi_{ABC}(\theta, x|y)k[(\theta', x')|(\theta, x)]\alpha[(\theta, x), (\theta', x')] \\
 &= \pi_{ABC}(\theta, x|y)k(\theta'|\theta)\pi(x'|\theta') \\
 &\quad \times \frac{\pi_\epsilon(y|x', \theta')\pi(\theta')k(\theta|\theta')}{\pi_\epsilon(y|x, \theta)\pi(\theta)k(\theta'|\theta)} \\
 &= \pi_\epsilon(y|x, \theta)\pi(\theta)\pi(x|\theta)k(\theta'|\theta)\pi(x'|\theta') \\
 &\quad \times \frac{\pi_\epsilon(y|x', \theta')\pi(\theta')k(\theta|\theta')}{\pi_\epsilon(y|x, \theta)\pi(\theta)k(\theta'|\theta)} \\
 &= \pi_\epsilon(y|x', \theta')\pi(\theta')\pi(x'|\theta') \underbrace{k(\theta|\theta')\pi(x|\theta)}_{k[(\theta, x)|(\theta', x')]} \\
 &= \pi_{ABC}(\theta', x'|y)p[(\theta, x)|(\theta', x')] \quad (۱۶.۲)
 \end{aligned}$$

بنابراین شرط مذکور برقرار است.

در عمل تولید نمونه‌های (θ_i, x_i) ، برای $i \geq 0$ ، در الگوریتم MCMC-ABC با استفاده از آماره‌های خلاصه‌شده خیلی راحت‌تر و کاراتر است. این کار باعث کاهش هزینه و زمان می‌گردد، به‌ویژه زمانی که مجموعه داده کمکی x_i بزرگ و پیچیده باشد. یک ویژگی جالب دیگر این روش نمونه‌گیری آن است که نرخ پذیرش آن به‌طور مستقیم به مقدار واقعی تابع درستنمایی $\pi(y|\theta')$ و بردار پیشنهادی θ' وابسته است. بدیهی است زمانی که از هسته یکنواخت در تابع تصمیم (۱۵.۲) استفاده می‌کنیم، حرکت به سمت (θ', x') زمانی پذیرفته می‌شود که $\rho(s(x), s(y)) \leq \epsilon$ و این حالت با احتمالی متناسب با نسبت درستنمایی اتفاق می‌افتد (سیزن و همکاران، ۲۰۰۷).

برای سطح تحمل کوچک ϵ ، نرخ پذیرش کاهش می‌یابد. این حالت به‌ویژه در دم‌های توزیع نمود بیشتری دارد. در نتیجه، زنجیر در نواحی پسینی کم‌چگال به ندرت قرار می‌گیرد. در این مواقع الگوریتم MCMC-ABC نرخ پذیرش بیشتری نسبت به الگوریتم RS-ABC دارد. دو حالت خاص را می‌توان ذکر کرد:

۱. اگر $k(\theta|\theta') = k(\theta'|\theta)$ ، آن‌گاه احتمال پذیرش الگوریتم ۶.۲ فقط به توزیع پیشین وابسته است.

۲. اگر $k(\cdot|\cdot)$ بازگشتی باشد (یعنی برای $\theta \neq \theta'$ ، $\pi(\theta)k(\theta'|\theta) = \pi(\theta')k(\theta|\theta')$)، آن‌گاه احتمال پذیرش برابر یک می‌شود و این یعنی پذیرش تمام نمونه پیشنهادی.

با توجه به مطالب بیان‌شده، با فرض ارگودیک بودن زنجیر (که در بالا برقراری آن را توسط شرط تعادل دقیق نشان دادیم)، می‌توان از نمونه‌های شبیه‌سازی‌شده برای تقریب پسین $\pi(\theta, x|y)$ استفاده کرد. مشابه روش‌های نمونه‌گیری MCMC، نمونه‌های آغازین تولیدشده توسط الگوریتم MCMC-ABC و همچنین الگوریتم RS-ABC (جداول ۶.۲ و ۴.۲) را می‌توان برای انجام استنباط کنار گذاشت (که مفهومی مشابه داغیدن^{۱۱} در روش‌های MCMC است)، زیرا زنجیر مارکوف حالت اولیه را فراموش می‌کند. در نتیجه با حذف مراحل اولیه، هزینه محاسبات کاهش پیدا می‌کند.

^{۱۱}Burn-in

داشتن تقریب ABC خوب به پارامترهای تنظیم‌کننده (آماره خلاصه شده $S(\cdot)$ ، فاصله $\rho(\cdot)$ و سطح تحمل ϵ) بستگی دارد. موثرترین و مهم‌ترین عامل تقریب، سطح تحمل ϵ است. به‌ویژه زمانی که ϵ به سمت صفر می‌رود، الگوریتم ABC دقیق‌تر می‌شود.

تابع چگالی پسین واقعی، $\pi(\theta|y)$ ، را می‌توان به‌طور مجزا تقریب زد؛ به‌طوری که $\pi(\theta|y)$ را با $\pi(\theta|s_y)$ و $\pi(\theta|s_y)$ را با $\pi_{ABC}(\theta|s_y)$ تقریب زد. در این تقریب زدن سه منبع خطا وارد مساله می‌شوند: خطای مونت کارلویی در تقریب پسین $\pi_{ABC}(\theta|s_y)$ ، خطای تقریب پسین ABC در تقریب پسین $\pi(\theta|s_y)$ و خطای تقریب $\pi(\theta|s_y)$ در تقریب $\pi(\theta|y)$. تشریح این مبحث به فصل ۳ موکول می‌شود.

۴.۲ عیوب و راه‌حل‌ها

استفاده از الگوریتم و روش‌های مختلف ABC، در مواردی که محاسبه تابع درست‌نمایی خیلی سخت یا ناممکن است، برای به‌دست آوردن تقریبی برای پسین واقعی $\pi(\theta|y)$ است. اگر مشاهدات y دارای بعد کوچکی باشد و $S(\cdot)$ تابع همانی باشد، یا آماره خلاصه شده با بعد کوچک برای داده‌ها وجود داشته باشد، این تقریب امکان پذیر است. در غیر این صورت، وقتی بعد داده‌ها یا آماره خلاصه شده بالا باشد، احتمال پذیرش نمونه‌های پیشنهادی کاهش می‌یابد و نمونه‌های بیشتری رد خواهند شد. این نتیجه به معنی عدم کارایی در محاسبات است.

از دیدگاه نظری، در این حالت با انتخاب ϵ به اندازه کافی کوچک، می‌توان خطای تقریب ABC را کنترل کرد. اما در عمل این توانایی چندان واضح نیست. بحث و استدلالی که برای حل داده‌های با بعد بالا مطرح است، استفاده از آماره‌های خلاصه شده‌ای است که تقریبی برای آماره‌های بسنده باشند. البته تعریف چنین آماره‌هایی چندان راحت نیست و مهم‌تر این‌که ساخت یک روش کلی برای پیدا کردن چنین آماره‌هایی، بیشتر اوقات سخت و حتی ناممکن است. یک رویکرد کلی پذیرفته شده (فرنهد و پرانگل، ۲۰۱۰؛ بلام، ۲۰۱۰؛ سیسلری و همکاران، ۲۰۱۰)، استفاده از میانگین پسین پارامترها به عنوان آماره خلاصه شده بهینه است. البته محاسبه میانگین پسین خود نیاز به محاسبه پسین $\pi(\theta|y)$ دارد، که هدف اصلی بحث است. یکی از راه‌حل‌ها، استفاده از روش‌های شبیه‌سازی مطرح شده در یک طرح اولیه و به‌دست آوردن برآوردی از آماره خلاصه شده بهینه است.

فصل ۳

چارچوب نظری روش‌های ABC

مقدمه

همان‌طور که در فصل قبل گفته شد، روش‌های ABC باید نسبت به خطایی که در تقریب توزیع پسین ایجاد می‌شود، تعادلی برقرار کنند، به طوری که تابع چگالی پسین $\pi_{ABC}(\theta|s_y)$ به چگالی پسین واقعی $\pi(\theta|y)$ نزدیک باشد، یا حداقل تقریب $\pi(\theta|\rho(s_x - s_y) \leq \epsilon)$ به $\pi(\theta|y)$ نزدیک باشد. در حالی که آماره خلاصه شده $S(\cdot)$ تابعی همانی از مشاهدات یا تقریبی از آماره بسنده باشد، می‌توان به تقریبی نزدیک به واقعیت امیدوار بود. در این فصل، به طور نظری، به برخی از راهکارهای کاهش خطای تقریب ABC می‌پردازیم. این راهکارها فقط برای حالتی که بعد $S(\cdot)$ کوچک باشد، امکان‌پذیرند. در غیر این صورت خطای مونت کارلوی تقریب، زیاد خواهد بود.

یکی از راهکارهایی که مطرح است، مبتنی بر تضعیف کردن شرایط حاکم بر مساله است، به طوری که $\pi_{ABC}(\theta|s_y)$ شبیه $\pi(\theta|y)$ گردد. با این راهکار، می‌توان برای π_{ABC} تقریب بسیار خوبی را، با توجه به دقت برآورد پارامتر مرکزی مدل، به دست آورد. برای این منظور، از یک ویژگی که کالبدن^۱ نامیده می‌شود، استفاده می‌کنیم. یکی از الگوریتم‌های ABC که به خوبی کالیبره شده است (روت و صادق، ۲۰۱۳)، الگوریتم ABC نوفه‌دار^۲ نامیده می‌شود. این الگوریتم، یک توزیع پسین ABC را نتیجه می‌دهد که هم ویژگی استنباطی خوبی دارد و هم می‌تواند روش‌های مونت کارلوی مطرح شده در فصل ۲ را به خوبی به کار ببرد، تا دقت و کالبدن الگوریتم را بررسی کند (جنتینگ، ۲۰۰۸).

بر اساس بررسی‌های انجام شده روی معیارهای دقت و کالبدن (بیومونت، ۲۰۰۲)، ثابت شده است که آماره خلاصه شده مورد نیاز همان میانگین پسین پارامترها، $E(\theta|y)$ ، است. این آماره بعد کوچکی دارد و خطای مونت کارلو را نسبتاً پایین نگه می‌دارد. در عمل، هنگامی که میانگین پسین پارامترها در دسترس نیست، با استفاده از برآورد آماره‌های بسنده به آماره خلاصه شده مورد نظر دست می‌یابیم. فرنیهد و پرائنگل (۲۰۱۱) برای حل مشکل انتخاب آماره‌های خلاصه شده، روش ABC نیمه خودکار^۳ را مطرح

^۳Semi-automatic ABC

^۱Calibration

^۲Noisy ABC

کردند. در حقیقت ABC نیمه خودکار چارچوبی را فراهم می‌کند که با استفاده از یک طرح شبیه‌سازی آزمایشی اولیه^۴، بتوان آماره خلاصه‌شده بهینه را، که همان میانگین توزیع پسین پارامترهاست، برآورد کرد.

۱.۳ کالبدن و ABC نوفه‌دار

در این بخش، به معرفی مفهوم کالبدن می‌پردازیم، که یک ویژگی مورد نیاز برای تقریب مناسب توزیع پسین می‌باشد. نشان می‌دهیم که، به‌طور معمول، توزیع هدف الگوریتم‌های ABC در فصل ۲، حتی برای مقادیر کوچک پهنای نوار ϵ ، کالیبره نیست. با یک اصلاح ساده که ABC نوفه‌دار نامیده می‌شود، می‌توان به توزیع هدفی دست یافت که کالیبره شده است (دین و همکاران، ۲۰۱۱).

کالبدن، پیشامدهایی را که با احتمال مثلاً q رخ می‌دهند را طبق نمونه‌گیری مکرر تعیین می‌کند. این نمونه‌گیری مکرر بر اساس خطایی که به مدل وارد می‌شود، انجام می‌گیرد. به‌عنوان مثال، توزیع پیشین کالیبره شده است. در ادامه کالبدن را به‌طور رسمی تعریف می‌کنیم و با یک مثال نشان می‌دهیم که توزیع پسین (۵.۲)، که به عنوان ABC استاندارد معرفی شد، کالیبره نیست. الگوریتمی جانشین که کالیبره شده است، الگوریتم ABC نوفه‌دار است. این نوع ABC طبق پسین ABC استاندارد اما بر اساس آماره‌های خلاصه‌شده تصادفی ارزیابی می‌شود؛ یعنی به‌جای آماره‌های خلاصه‌شده معمول از آماره‌های خلاصه‌شده که با یک نوفه جمع شده‌اند، استفاده می‌شود.

۱.۱.۳ کالبدن چیست؟

مدل احتمالی $\pi(y|\theta)$ را با داده y و پارامتر θ در نظر بگیرید. تقریبی که برای توزیع پسین این مدل، تحت توزیع پیشین $\pi(\theta)$ برای پارامترها، تعریف می‌شود را با (g, s) نشان می‌دهیم، به‌طوری که $s = S(y)$ مقدار مشاهده‌شده آماره خلاصه‌شده بر حسب y و $g = g(\theta; s)$ تابع چگالی پسین s است. توجه کنید تابع چگالی آماره خلاصه‌شده S ممکن است $\pi(s|y)$ باشد اما تابع چگالی پسین آن، یعنی $g(\theta; s)$ ، لزوماً با $\pi(\theta|s)$ یکی نیست. بنابراین، بر اساس این ایده، اگر مشاهده y را داشته باشیم، آن‌گاه یک تحقق از s, S خواهد بود و یک تقریب برای تابع چگالی پسین، $g(\theta; s)$ است.

با این نمادگذاری، تابع چگالی پسین ABC استاندارد را می‌توان با $g(\theta; s) = \pi_{ABC}(\theta|s)$ بیان کرد. الگوریتم ABC نوفه‌دار از تابع چگالی پسین ABC، به‌عنوان $g(\theta; s)$ ، استفاده می‌کند با این تفاوت که آماره خلاصه‌شده به‌کار گرفته‌شده در آن به صورت

$$s = S(y) + \epsilon x, \quad (1.3)$$

تعریف می‌شود، که در آن x از چگالی با هسته $\mathcal{K}(\cdot)$ تولید می‌شود و ϵ جمله نوفه است که می‌تواند دارای توزیع نرمال باشد. اکنون یک تقریب کالیبره را تعریف می‌کنیم.

^۴Pilot design

تعریف ۱.۱.۳. (پرانگل، ۲۰۱۱) فرض کنید پیشامد \mathcal{A} یک زیرمجموعه از فضای پارامتر باشد و $q \in [0, 1]$ برای $\lambda > 0$ تعریف می‌کنیم

$$E_{q,\lambda}(\mathcal{A}) = \left\{ s : \left| \int_{\theta \in \mathcal{A}} g(\theta; s) d\theta - q \right| \leq \lambda \right\}. \quad (2.3)$$

تقریب (g, s) کالیبره است، اگر برای هر \mathcal{A} و q ، $P(s \in E_{q,\lambda}(\mathcal{A})) > 0$ باشد، و برای هر $\lambda > 0$ به اندازه کافی کوچک، داشته باشیم

$$\lim_{\lambda \rightarrow 0} P(\theta \in \mathcal{A} | s \in E_{q,\lambda}(\mathcal{A})) = q. \quad (3.3)$$

عبارت (۳.۳) بیانگر آن است که طبق نمونه‌گیری مکرر پارامترها از توزیع پیشین و داده‌ها و آماره‌های خلاصه‌شده از مدل، اگر فراوانی نسبی پیشامدی برابر q باشد، آن‌گاه تقریب $g(\theta; s)$ نیز احتمال q را به آن پیشامد نسبت می‌دهد. دو عبارت (۲.۳) و (۳.۳) تکمیل‌کننده هم و معادل یکدیگر می‌باشند. یعنی اگر

$$\left| \int_{\theta \in \mathcal{A}} g(\theta; s) d\theta - q \right| \leq \lambda,$$

آن‌گاه

$$-\lambda \leq \int_{\theta \in \mathcal{A}} g(\theta; s) d\theta - q \leq \lambda,$$

که می‌توان نوشت

$$\int_{\theta \in \mathcal{A}} g(\theta; s) d\theta = q \pm \lambda.$$

بنابراین

$$\lim_{\lambda \rightarrow 0} \int_{\theta \in \mathcal{A}} g(\theta; s) d\theta = q.$$

نتیجه کالبدن، این است که تقریب پسین g تقریباً بیانگر عدم قطعیت واقعی موجود در پارامترها است.

۲.۱.۳ الگوریتم ABC نوفه‌دار

برای به‌دست آوردن الگوریتم نمونه‌گیری ABC نوفه‌دار، کافیت تا یک مرحله به الگوریتم‌های نمونه‌گیری ۵.۲ و ۶.۲ اضافه کنیم. این مرحله در جدول ۱.۳ بیان شده است. در این الگوریتم چنان‌چه ϵ به صفر میل کند، آن‌گاه ABC نوفه‌دار با ABC استاندارد معادل می‌شود (فرنهید و پرانگل، ۲۰۱۰). در ادامه دو قضیه در مورد ویژگی کالبدن الگوریتم‌های ABC و ABC نوفه‌دار مطرح می‌کنیم که نشان می‌دهد الگوریتم نمونه‌گیری ABC نوفه‌دار کالیبره است و تحت شرایطی الگوریتم نمونه‌گیری ABC استاندارد نیز می‌تواند کالیبره باشد.

جدول ۱۰.۳: افزودن مرحله آماره خلاصه‌شده تصادفی در الگوریتم ABC نوفه‌دار

ABC نوفه‌دار
گام زیر به الگوریتم‌های نمونه‌گیری ۵.۲ و ۶.۲ اضافه می‌گردد:
x را از $\mathcal{K}(\cdot)$ شبیه‌سازی کن و قرار بده $s = S(y) + \epsilon x$.

قضیه ۲.۱.۳. (فیرنهد و پرانگل، ۲۰۱۰) اگر $g(\theta; s) = \pi(\theta|s)$ ، آن‌گاه $g(\theta; s)$ کالیبره است.

برهان. فرض کنید $g(\theta; s) = \pi(\theta|s)$ و A و q ثابت باشند، به گونه‌ای که برای λ به اندازه کافی کوچک، $P(s \in E_{q,\lambda}(A)) > 0$. برای $s \in E_{q,\lambda}(A)$ داریم

$$\int_{\theta \in A} \pi(\theta|s) d\theta = q + \delta_s(\lambda) \quad (۴.۳)$$

به طوری که $\delta_s(\lambda) < \lambda$. با استفاده از روابط (۵.۲) و (۶.۲)، و همچنین قرار دادن $\pi(s|y, \theta)$ به جای $\mathcal{K}((s_x - s_y)/\epsilon)$ و سپس انتگرال‌گیری از طرفین (۴.۳) نسبت به $s \in E_{q,\lambda}(A)$ و D (تکیه‌گاه y) می‌توان نوشت

$$\int_{\theta \in A} \pi(\theta) \int_{s \in E_{q,\lambda}(A)} \pi(y|\theta) \pi(s|y, \theta) dy ds d\theta = \int_{s \in E_{q,\lambda}(A)} (q + \delta_s(\lambda)) dy ds.$$

در نتیجه

$$\int_{\theta \in A} \pi(\theta) \int_{s \in E_{q,\lambda}(A)} \frac{\pi(y, \theta) \pi(s, y, \theta)}{\pi(\theta) \pi(y, \theta)} dy ds d\theta = (q + \delta(\lambda)) \int_{s \in E_{q,\lambda}(A)} dy ds.$$

و بنابراین

$$\int_{\theta \in A, s \in E_{q,\lambda}(A)} \pi(\theta, y, s) d\theta dy ds = (q + \delta(\lambda)) \int_{s \in E_{q,\lambda}(A)} dy ds = (q + \delta(\lambda)) P(s \in E_{q,\lambda}(A)),$$

که در آن $|\delta(\lambda)| \leq \lambda$. در نتیجه با توجه به تعریف ۱.۱.۳ و میل دادن λ به سمت صفر، کالیبره بودن $g(\theta; s)$ اثبات می‌شود. \square

برهان دیگری برای قضیه ۲.۱.۳ در قضیه ۱ فیرنهد و پرانگل (۲۰۱۰)، به صورت جبری، ارایه شده است که نشان می‌دهد الگوریتم ABC نوفه‌دار کالیبره است. بنابراین تا این جا می‌توان گفت توزیع‌های پیشین، پسین و ABC نوفه‌دار، کالیبره هستند. به منظور این‌که شرایط مورد نیاز برای کالیبره بودن ABC نوفه‌دار حفظ گردد، با استفاده از دو رابطه (۵.۲) و (۶.۲)، داریم

$$\pi_{ABC}(\theta|s) \propto \pi(\theta) \int \pi(s|\theta) \mathcal{K}((s_x - s_y)/\epsilon) ds, \quad (۵.۳)$$

که انتگرال رابطه (۵.۳) را می‌توان به عنوان پیچشی از تابع چگالی پسین واقعی s و هسته $\mathcal{K}(\cdot)$ تعبیر کرد به طوری که s توسط رابطه (۲.۳) تولید شده باشد.

در همین زمینه، ایده مرتبطی توسط ویلکینسون (۲۰۰۸) مطرح شد. او نشان داد توزیع پسین ABC استاندارد، تحت یک پذیره مناسب برای خطای مدل، با توزیع پسین واقعی معادل است. در واقع، ویلکینسون (۲۰۰۸) پیشنهاد کرد تا تقریب پسین حاصل از نمونه‌گیری ABC، که با ضعیف کردن شرایط پذیرش به دست می‌آید، را با یک تقریب کنترل‌شده که از پیچش توزیع هدف و یک تابع هسته اختیاری نتیجه می‌شود، جایگزین کنیم. در این حالت توزیع هدف متناظر ABC به صورت زیر است:

$$\pi_\epsilon(\theta, x|y) = \frac{\pi(\theta)\pi(x|\theta)\mathcal{K}_\epsilon(y-x)}{\int \pi(\theta)\pi(x|\theta)\mathcal{K}_\epsilon(y-x)dsd\theta}, \quad (۶.۳)$$

که در آن $\mathcal{K}_\epsilon(\cdot)$ تابع هسته پارامتری با پهنای نوار ϵ است. این دیدگاه از این منظر که خروجی الگوریتم به طور کامل کنترل می‌گردد، جالب است. دلیل کنترل‌پذیر بودن خروجی نیز درجه آزادی است که از انتخاب تابع هسته نتیجه می‌شود. ویلکاکسون (۲۰۰۸) یک مرحله با ارزش را به الگوریتم ABC افزود، به گونه‌ای که اگر مدل شامل خطا گردد، آنگاه توزیع حاصل از خطا در $\mathcal{K}_\epsilon(\cdot)$ ، منجر به یک الگوریتم ABC می‌شود که شبیه‌سازی دقیقی از متغیر خطا در پسین را شامل خواهد شد. در عمل روش ویلکینسون نیاز به یک اصلاح در الگوریتم ABC استاندارد دارد که با توجه به در نظر گرفتن هسته در شبیه‌سازی x ، الگوریتم جدید شامل یک گام بیشتر در مرحله پذیرش یا رد شده و برای پیچش حاصل از هسته $\mathcal{K}_\epsilon(\cdot)$ کران بالایی را ارائه می‌دهد.

قضیه زیر کرانی برای کالبدن نادرست، تحت ضعیف شدن شرایط، ارائه می‌دهد و نشان می‌دهد که ABC استاندارد برای ϵ کوچک تقریباً کالیبره است. این قضیه، کالیبره بودن الگوریتم ABC استاندارد را تحت دو شرط

(۱) میانگین $\mathcal{K}(\cdot)$ صفر است و

(۲) همه گشتاورهای $\mathcal{K}(\cdot)$ کراندار هستند

برای تابع هسته، بررسی می‌کند.

قضیه ۳.۱.۳. (پرانگل، ۲۰۱۱) فرض کنید برای زیرمجموعه‌ای از فضای پارامتر \mathcal{A} و احتمال q و برای هر $\lambda > 0$ به اندازه کافی کوچک،

$$P(s \in E_{q,\lambda}(\mathcal{A})) > 0 \quad (۳)$$

$$P(\theta \in \mathcal{A}, s+x \in E_{q,\lambda}(\mathcal{A})) \text{ و } P(s+x \in E_{q,\lambda}(\mathcal{A})) \quad (۴)$$

در این صورت، برای هر s

$$\lim_{\lambda \rightarrow 0} P(\theta \in \mathcal{A} | s \in E_{q,\lambda}(\mathcal{A})) = q + O(\epsilon^2).$$

۲.۳ سه سطح تقریب و سه خطا برای برآورد پسین پارامترها

در فصل دوم اشاره کردیم که در تقریب توزیع پسین با استفاده از روش‌های ABC سه خطا رخ می‌دهد. اولین خطا به سبب خطای مونت کارلویی است که در تقریب توزیع پسین رخ می‌دهد. دومین خطا به علت استفاده از آماره خلاصه‌شده $S(\cdot)$ به جای داده‌های اصلی می‌باشد؛ یعنی زمانی که تابع چگالی پسین $\pi(\theta|y)$ با $\pi(\theta|s_y)$ تقریب زده می‌شود. سومین خطا ناشی از استفاده از هسته و نرم آماری در تقریب توزیع پسین آماره خلاصه‌شده با استفاده از ABC است؛ یعنی تقریب تابع چگالی پسین $\pi(\theta|s_y)$ با $\pi(\theta|\rho(s_x, s_y) \leq \epsilon)$ در این بخش ابتدا، در زیربخش ۱.۲.۳، خطای مونت کارلوی الگوریتم ABC را بررسی می‌کنیم، به‌ویژه نشان می‌دهیم که این خطا نسبت به انتخاب آماره‌های خلاصه‌شده و سطح تحمل ϵ تاثیرپذیر است. سپس در زیربخش ۲.۲.۳ خطای تقریب پسین ناشی از استفاده از آماره‌های خلاصه‌شده را مورد بحث قرار می‌دهیم. در این بخش، به تعریف معیار دقت می‌پردازیم و در پی آن از میانگین پسین ABC به عنوان آماره خلاصه‌شده بهینه استفاده می‌کنیم. در انتها نیز در زیربخش ۳.۲.۳ به بررسی انتخاب خطای آستانه تحمل (یا هسته) و نرم در تقریب پسین می‌پردازیم.

۱.۲.۳ خطای مونت کارلو ABC

در این بخش به بررسی خطای مونت کارلوی الگوریتم ABC می‌پردازیم. این خطا را برای هر سه نوع الگوریتم ABC که در فصل دوم معرفی کردیم، بررسی خواهیم کرد. نشان خواهیم داد که برای هر سه الگوریتم، خطای مونت کارلو به عوامل مشابه وابسته و حساس هستند. به عنوان مثال، نمونه‌گیری RS-ABC به نرخ پذیرش الگوریتم وابسته است و رفتار جانبی آن نسبت به ϵ کوچک حساس است.

نمونه‌گیری RS-ABC

با توجه به بخش ۲.۳.۲، تقریب $\pi(\theta|s_y)$ با خروجی $(\theta_i, w_i)_{1 \leq i \leq T}$ از الگوریتم نمونه‌گیری نقاط مهم (که نمونه‌گیری رد حالت خاصی از آن است)، را می‌توان به‌عنوان برآورد تابع چگالی پسین ABC در نظر گرفت و برای برآورد عددی تابعی مثل $h(\theta)$ به شکل زیر استفاده کرد:

$$\hat{h}_{ABC}(\theta) = \frac{\sum_{i=1}^T h(\theta_i) w_i}{\sum_{i=1}^T w_i} \quad (7.3)$$

در نمونه‌گیری رد الگوریتم ۵.۲، w_i ها صفر یا یک هستند. واریانس جانبی این برآوردگر برای T های بزرگ به‌صورت زیر است:

$$\frac{Var(h(\theta)|s_y)}{T_{acc}}, \quad (8.3)$$

که در آن، صورت کسر، واریانس پسین ABC برای برآوردگر $h(\theta)$ و مخرج کسر

$$T_{acc} = T \int \pi(\theta|s_y) \pi(\theta) d\theta,$$

است، که تعداد تکرارهای پذیرفته شده مورد انتظار است (به خاطر داشته باشید که $\pi(\theta|s_y)$ تقریب ABC با درستی (۶.۲) است). از (۸.۳) واضح است که واریانس برآوردگر حاصل از تقریب ABC به نرخ پذیرش الگوریتم وابسته است. هر چه نرخ پذیرش بزرگتر باشد، دقت برآوردگر بیشتر خواهد بود.

نمونه گیری IS-ABC

محاسبه واریانس برآوردگر میانگین پسین بر اساس نمونه گیری نقاط مهم به دلیل وابستگی بین وزن های w_i و مقادیر $h(\theta)$ پیچیده است (لیو، ۱۹۹۶ و پرانگل، ۲۰۱۱ را ببینید). روش معمول که دقت نمونه گیری نقاط مهم را تعیین می کند، استفاده از حجم نمونه موثر^۵ (ESS) (گلمن و منگ، ۱۹۹۸) است. اگر وزن های نرمال شده دارای میانگین برابر یک باشند، آن گاه T_{ess} ، ESS، برابر است با T تقسیم بر میانگین توان های دوم وزن ها (لیو، ۱۹۹۶). نشان داده شده است برای اکثر توابع $h(\theta)$ واریانس برآوردگر (۷.۳)، تقریباً برابر (۸.۳) است، البته با جایگزین کردن T_{ess} به جای T_{acc} . لیو (۱۹۹۶) از وزن هایی استفاده کرد که نشان دهد وزن های تصادفی الگوریتم نمونه گیری ۵.۲ همین گزاره را نتیجه خواهد داد.

در الگوریتم ۵.۲، از وزن های $w(\theta) = \mathcal{X}\pi(\theta)/g(\theta)$ استفاده می شود، که در آن $g(\cdot)$ توزیع پیشنهادی و \mathcal{X} یک متغیر تصادفی برنولی مستقل با $p(\mathcal{X} = 1) = \mathcal{K}((s_x - s_y)/\epsilon)$ است. به سادگی می توان نتیجه گرفت، برای عدد صحیح $n > 0$ ، $E(w^n|\theta) = E(\mathcal{X}|\theta) \left(\frac{\pi(\theta)}{g(\theta)}\right)^n$. دقت کنید که این امید ریاضی روی مقادیر s ، برای داده های شبیه سازی شده محاسبه می شود و $E(\mathcal{X}|\theta) = \pi(\theta|s_y)$ برای این الگوریتم، ESS به صورت

$$T_{eff} = \frac{TE_g^\mathcal{X}(w)}{E_g(w^2)}, \quad (9.3)$$

محاسبه می شود که در آن، زیر نویس g برای تاکید آن است که امید ریاضی تحت تابع چگالی پیشنهادی محاسبه می شود. همچنین

$$\begin{aligned} E_g(w) &= E_g(E(w|\theta)) \\ &= \int \pi(\theta|s_y)\pi(\theta)d\theta. \end{aligned} \quad (10.3)$$

و

$$\begin{aligned} E_g(w^2) &= E_g(E(w^2|\theta)) \\ &= \int \frac{\pi(\theta)}{g(\theta)}\pi(\theta|s_y)\pi(\theta)d\theta. \end{aligned} \quad (11.3)$$

اکنون به بررسی توزیع پیشنهادی بهینه با توجه به ماکسیمم T_{eff}/T می پردازیم. چون مطلوب است واریانس مونت کارلو کاهش یابد، باید $E_g^\mathcal{X}(w)$ که متناسب است با $\int \frac{q(\theta)}{g(\theta)}q(\theta)d\theta$ ، می نیمم شود،

به طوری که $q(\theta)$ چگالی متناسب با $\pi(\theta|s_y)^{1/2}\pi(\theta)$ است. با توجه به ليو (۱۹۹۶) و این که وزن‌ها دارای میانگین ۱ هستند، انتگرال $\int \frac{q(\theta)}{g(\theta)}q(\theta)d\theta$ را می‌توان، با انتخاب $g(\theta)$ ، حول ۱ کراندار کرد؛ زیرا می‌توان آن را به صورت

$$E_g(w^2) = \int \frac{q^2(\theta)}{g^2(\theta)}g(\theta)d\theta \quad (12.3)$$

$$= 1 + Var\left(\frac{q(\theta)}{g(\theta)}\right), \quad (13.3)$$

نوشت. برای به دست آوردن توزیع پیشنهادی بهینه کفایت قرار دهیم $g(\theta) = q(\theta)$. این انتخاب برای توزیع پیشنهادی، واریانس عبارت (۱۳.۳) را صفر و کران مورد نظر را می‌نیم می‌کند. بر اساس این توزیع پیشنهادی داریم

$$\begin{aligned} E_g(w^2) &= E_g(E(w^2|\theta)) \\ &= E_g\left(E(\mathcal{X}|\theta)\left(\frac{\pi(\theta)}{g(\theta)}\right)^2\right) \\ &= E_g\left(\pi(\theta|s_y)\left(\frac{\pi(\theta)}{g(\theta)}\right)^2\right) \\ &= \int \pi(\theta|s_y)\left(\frac{\pi(\theta)}{g(\theta)}\right)^2g(\theta)d\theta \\ &= \int \pi(\theta|s_y)\frac{(\pi(\theta))^2}{g(\theta)}d\theta \\ &= \int \pi(\theta|s_y)\frac{(\pi(\theta))^2}{\pi(\theta)(\pi(\theta|s_y))^{1/2}}d\theta \\ &= \int \pi(\theta|s_y)^{1/2}\pi(\theta)d\theta \\ &= E_\pi(\pi(\theta|s_y)^{1/2}). \end{aligned} \quad (14.3)$$

همچنین

$$\begin{aligned} E_g(w) &= E_g\left(\pi(\theta|s_y)\frac{\pi(\theta)}{g(\theta)}\right) \\ &= \int \pi(\theta|s_y)\frac{\pi(\theta)}{g(\theta)}g(\theta)d\theta \\ &= E_\pi(\pi(\theta|s_y)) \\ &= Var_\pi(\pi(\theta|s_y)^{1/2}) + E_\pi^2(\pi(\theta|s_y)^{1/2}). \end{aligned} \quad (15.3)$$

با توجه به دو عبارت (۱۴.۳) و (۱۵.۳)، ESS به صورت زیر تبدیل می‌شود:

$$\begin{aligned} T_{ess} &= T_{acc} \left(\frac{(Var_\pi(\pi(\theta|s_y)^{1/2}) + E_\pi^2(\pi(\theta|s_y)^{1/2}))^2}{E_\pi(\pi(\theta|s_y)^{1/2})} \right) \\ &= \left(\frac{Var_\pi^2(\pi(\theta|s_y)^{1/2})}{E_\pi(\pi(\theta|s_y)^{1/2})} + E_\pi^2(\pi(\theta|s_y)^{1/2}) + 2Var_\pi(\pi(\theta|s_y)^{1/2})E_\pi(\pi(\theta|s_y)^{1/2}) \right). \end{aligned} \quad (16.3)$$

مشاهده می‌کنید که حجم نمونه موثر به T_{acc} وابسته است. بدیهی است که $T_{ess} \geq T_{acc}$. بنابراین واضح است که استفاده از نمونه‌گیری IS-ABC می‌تواند به مراتب استنباط‌های دقیق‌تری را (از منظر خطای مونت کارلو) نسبت به نمونه‌گیری RS-ABC فراهم کند. البته T_{ess} خیلی کوچک، منجر به پدید آمدن وزن‌های نامتعادل^۶ شده و باعث می‌شود نتایج استنباط غیر قابل اطمینان شوند. به این معنی که واریانس مونت کارلو افزایش می‌یابد (افزایش خطا) و نمونه‌های پذیرفته‌شده نیز افزایش می‌یابند که باعث کاهش بیش از حد سطح تحمل ϵ می‌گردد. زمانی که $\pi(\theta|s_y)$ به θ وابسته نباشد، $T_{ess} = T_{acc}$.

MCMC-ABC نمونه‌گیری

تحلیل خطای مونت کارلو در الگوریتم MCMC-ABC سخت‌تر است. با این حال به بررسی هسته پیشنهادی $k(\cdot|\cdot)$ می‌پردازیم. سعی بر این است هسته پیشنهادی به شکلی انتخاب شود تا تابع انتقال را ثابت نگه دارد. خطای مونت کارلو در درجه اول توسط میانگین احتمال پذیرش کنترل می‌گردد. برای سادگی کار فرض کنید \mathcal{K} هسته یکنواخت باشد و $k(\cdot|\cdot)$ طوری انتخاب شود که توزیع مانای آن $\pi(\theta)$ باشد. به عبارت دیگر، در مرحله i ام نمونه‌گیری $1 \approx \pi(\theta)k(\theta^{i-1}|\theta) / \{\pi(\theta^{i-1})k(\theta|\theta^{i-1})\}$ و بتوان این نسبت را نادیده گرفت. میانگین احتمال پذیرش به صورت زیر تعریف می‌شود (پرانگل، ۲۰۱۱):

$$\iint \pi_{ABC}(\theta|s_y)k(\theta'|\theta)\pi(\theta'|s_y)d\theta d\theta' = \left(\int \pi(\theta)\pi(\theta|s_y)d\theta \right) \times \iint \pi_{ABC}(\theta|s_y)k(\theta'|\theta) \frac{\pi_{ABC}(\theta'|s_y)}{\pi(\theta')} d\theta d\theta'. \quad (17.3)$$

این انتگرال در حقیقت بیان‌کننده میانگین مقادیر پیشنهادی و فعلی برای الگوریتم MCMC است و همچنین میانگین احتمال پذیرش برای θ' ای که از $\pi(\theta'|s_y)$ آمده است. سمت راست تساوی (۱۷.۳) با استفاده از (۵.۲) به دست آمده است. عبارت اول سمت راست (۱۷.۳)، بیان‌کننده میانگین احتمال پذیرش الگوریتم رد می‌باشد. دومین عبارت سمت راست، برای نمونه‌گیری مستقل از $k(\theta'|\theta)$ ، برابر یک می‌شود. اگر پسین ABC نسبت به پیشین به اوج خود (قله) برسد و هسته انتقال حرکت موضعی را پیشنهاد کند (مانند الگوریتم قدم زدن تصادفی)، آن‌گاه دومین عبارت سمت راست (۱۷.۳)، $1 \ll$ خواهد شد.

ویژگی‌های مجانبی

در هر سه الگوریتم معرفی‌شده، خطای مونت کارلو اکیدا به $\int \pi(\theta|s_y)\pi(\theta)d\theta$ وابسته است، که همان میانگین پذیرش الگوریتم رد می‌باشد. لم زیر برای سطح تحمل کوچک ϵ و در حالتی که آماره خلاصه‌شده پیوسته باشد، احتمال پذیرش را مشخص می‌کند. برای حالت گسسته، به‌طور مشابه، قابل اثبات است.

لم ۱۰.۲.۳. فرض کنید یکی از دو بند زیر برقرار باشند:

الف) تابع چگالی حاشیه‌ای آماره خلاصه‌شده، $\pi(s)$ ، در $S = s_y$ پیوسته و دامنه تابع هسته $\mathcal{K}(\cdot)$ متناهی باشد.

^۶Imbalanced weight

(ب) تابع چگالی $\pi(s)$ مشتق‌پذیر پیوسته باشد و برای هر i ، $|\partial\pi(s)/\partial s_i|$ از بالا کراندار و $\int |x_i|\mathcal{K}(x)dx$ کراندار باشد.

آن‌گاه وقتی $\epsilon \rightarrow 0$ ،

$$\int \pi(\theta|s_y)\pi(\theta)d\theta = \pi(s_y)\epsilon^d + o(\epsilon^d), \quad (18.3)$$

که در آن d بعد s_y است.

برهان. با توجه به (۵.۲)، (۶.۲) و با استفاده از تعریف $\pi(\theta|s_y)$ ، می‌توان نوشت

$$\begin{aligned} \int \pi(\theta|s_y)\pi(\theta)d\theta &= \iint \mathcal{K}((s_x - s_y)/\epsilon)\pi(s_x|\theta)\pi(\theta)d\theta ds_x \\ &= \int \epsilon^d \mathcal{K}(x) \int \pi(s_y + \epsilon x|\theta)\pi(\theta)d\theta dx \\ &= \int \epsilon^d \mathcal{K}(x) \int \frac{\pi(s_y + \epsilon x, \theta)}{\pi(\theta)}\pi(\theta)d\theta dx \\ &= \int \epsilon^d \mathcal{K}(x)\pi(s_y + \epsilon x)dx. \end{aligned}$$

دومین تساوی از انتگرال‌گیری روی θ و با تغییر متغیر $x = (s_x - s_y)/\epsilon$ حاصل شده است. بنابراین با توجه به نامساوی ینسن^۷ می‌توان نوشت

$$\left| \epsilon^{-d} \int \pi(\theta|s_y)\pi(\theta)d\theta - \pi(s_y) \right| \leq \int \mathcal{K}(x)|\pi(s_y + \epsilon x) - \pi(s_y)|dx.$$

این کران در هر دو شرط با $O(1)$ نشان داده شده است.

ابتدا شرط (الف) را بررسی می‌کنیم. ثابت c را طوری تعریف می‌کنیم که مقدار $|x|$ برای $\mathcal{K}(x) > 0$ ماکسیمم گردد. برای هر $\gamma > 0$ و با توجه به تعریف پیوستگی $\pi(s)$ در s_y ، $\delta > 0$ وجود دارد که اگر $|x| < \delta$ ، آن‌گاه $|\pi(s_y + x) - \pi(s_y)| < \gamma$. قرار می‌دهیم $\epsilon_\gamma = \delta/c$. بنابراین، برای هر $\epsilon < \epsilon_\gamma$ ، داریم

$$\int \mathcal{K}(x)|\pi(s_y + \epsilon x) - \pi(s_y)|dx \leq \gamma. \quad (19.3)$$

در این نامساوی به ازای هر x که $\mathcal{K}(x) > 0$ باشد، $\epsilon(x) < \delta$ است.

اکنون شرط (ب) را بررسی می‌کنیم. به کمک ویژگی مشتق‌پذیری پیوسته و قضیه تیلور

$$\pi(s_y + \epsilon x) = \pi(s_y) + \sum_i \epsilon x_i r_i(x),$$

که در آن جمله $r_i(x)$ باقیمانده بسط تیلور و تابعی است از $|\partial\pi(\cdot)/\partial s_i|$. بنابراین با جایگذاری بسط به دست آمده در عبارت (۱۹.۳) و با فرض $|r_i(x)| \leq R$ ، کران متناهی می‌گردد. یعنی

$$\int \mathcal{K}(x)|\pi(s_y + \epsilon x) - \pi(s_y)|dx \leq \epsilon R \sum_i \int |x_i|\mathcal{K}(x)dx = O(1).$$

□

^۷Jensen's inequality

این لم نشان می‌دهد که $S(\cdot)$ و ϵ روی خطای مونت کارلو تاثیر دارند. مقادیر کوچک ϵ^d ، خطای مونت کارلو را بزرگ می‌کنند. بنابراین ایده‌آل است که $S(\cdot)$ ، خلاصه‌ای با بعد کوچکی از داده‌ها باشد، یا برای پارامتر θ بسنده باشد. انتخاب ϵ روی دقت روش ABC در تقریب $\pi(\theta|s_y)$ و همچنین روی میانگین احتمال پذیرش (۱۸۰۳) و خطای مونت کارلو نیز تاثیر می‌گذارد. توجه کنید که لم ۱۰۲۰۳ را می‌توان به عنوان تاثیر بعد معرفی کرد که در بخش ۱۰۳۰۲ نیز بررسی شده است. در بخش ۳۰۲۰۳، انتخاب ϵ بهینه که خطا را می‌نیم کند، بررسی می‌شود. دسترسی به مشاهدات واقعی، این انگیزه را ایجاد می‌کند تا به جای تمرکز بر روی تقریب آماره‌های بسنده که معمولا دارای ابعاد بزرگ هستند، از برخی رویکردهای بهینه‌سازی، شامل دقت و صحت اندازه‌گیری، که می‌توانند با یک مجموعه از آماره‌های خلاصه‌شده به دست آیند، استفاده کنیم.

۲۰۲۰۳ خطای تقریب پسین در استفاده از آماره‌ها (غیربسنده)

همان‌طور که در ابتدای بخش ۲۰۳ اشاره شد، سطح دومی از خطای تقریب توزیع پسین، ناشی از خلاصه کردن محسوس داده‌ها در قالب استفاده از آماره‌های خلاصه‌شده‌ای است که غالبا بسنده نیستند. قسمتی از تحقیقات بر روی این موضوع و انتخاب آماره‌های خلاصه‌شده، توسط جویس و ماجروم (۲۰۰۸) و فرنهید و پرانگل (۲۰۱۲) صورت گرفته است. جویس و ماجروم (۲۰۰۸) اساس انتخاب را روی اندازه مربوط به مفهوم تقریب بسندگی قرار دادند. فرنهید و پرانگل (۲۰۱۲) از دیدگاه استنباطی و کالیبره بودن تقریب ABC برای برآورد توزیع هدف، به انتخاب آماره‌های خلاصه‌شده پرداختند. آن‌ها اساس کار خود را بر آماره خلاصه‌شده تصادفی که نسخه به‌خوبی کالیبره‌شده ABC است، بنیان نهادند. به عبارت دیگر الگوریتمی را پیشنهاد دادند که برآوردهای صحیحی (مناسبی) از چندک‌های توزیع هدف را ارائه می‌دهد. فرنهید و پرانگل (۲۰۱۲) دریافتند که امید ریاضی توزیع پسین پارامتر مورد علاقه، همان آماره خلاصه‌شده بهینه است. البته این انتخاب بهینه به انتخاب تابع زیان^۸ وابسته است. از این رو، معیار دیگری معرفی کردند که تحت عنوان معیار دقت در برآورد پسین پارامترها معرفی شد.

دقت و انتخاب آماره‌های خلاصه‌شده

کالبدن به خودی خود برای کاهش خطای تقریب پسین کافی نیست و باید یک معیار دیگر در سنجش ABC مورد استفاده قرار گیرد که خطای تقریب پسین را به حداقل برساند. این کمیت، همان معیار دقت است. توزیع پیشین کالیبره است اما کالیبره بودن دقتی از برآورد پارامتر را به ما ارائه نمی‌دهد (فرنهید و پرانگل، ۲۰۱۰). از طرفی، همیشه به دنبال داشتن ماکسیمم دقت برآورد در تقریب پسین ABC هستیم. برای بررسی دقت در برآورد پسین ناچار به بررسی توابع زیان در این زمینه هستیم. یک انتخاب طبیعی، تابع زیان توان دوم خطا^۹ است. فرض کنید θ مقدار واقعی پارامتر و $\hat{\theta}$ برآوردی برای θ باشد. رده‌ای از توابع زیان که در این زمینه استفاده می‌شود، به شکل زیر است:

$$L(\theta_0, \hat{\theta}; A) = (\theta_0 - \hat{\theta})^T A (\theta_0 - \hat{\theta}), \quad (20.3)$$

^۹Quadratic loss function

^۸ Loss function

به طوری که A یک ماتریس وزن معین مثبت^{۱۰} با بعد p است، که p معرف بعد پارامترها است. با توجه به دیدگاه استنباطی فیرنهد و پیرانگل (۲۰۱۲)، هدف می‌نیم کردن این تابع زیان است. زمانی که ϵ به سمت صفر میل می‌کند، نتایجی حاصل می‌گردد که به انتخاب بهینه آماره خلاصه شده $S(\cdot)$ منجر می‌شود. قضیه زیر نشان می‌دهد، برای هر انتخاب ماتریس وزنی رتبه کامل^{۱۱} A ، انتخاب آماره خلاصه شده بهینه، $s_y = E(\theta|y)$ است، که همان میانگین پسین واقعی است. این قضیه فقط برای روش‌های ABC استاندارد و ABC نوفه‌دار کاربرد دارد. در زیر منظور از $tr(B)$ ، اثر ماتریس B است. قضیه ۲.۲.۳. با توجه به تعریف ۲.۰.۳، فرض کنید ماتریس رتبه کامل A معین مثبت $p \times p$ باشد. همچنین فرض کنید ماتریس Σ ، ماتریس واریانس توزیع پسین واقعی θ باشد.

الف) برای $\hat{\theta} = E(\theta|y)$ ، می‌نیم تابع توان دوم خطا، $E(L(\theta, \hat{\theta}; A)|y)$ ، برابر است با $tr(A\Sigma)$.

ب) اگر $s_y = E(\theta|y)$ ، آنگاه چنانچه $\epsilon \rightarrow 0$ ، می‌نیم تابع زیان توان دوم خطا مبتنی بر پسین ABC، به ازای $\hat{\theta} = E_{ABC}(\theta|s_y)$ حاصل می‌شود. در نتیجه امید ریاضی تابع زیان (تابع مخاطره^{۱۲})، همان $tr(A\Sigma)$ است.

برهان. این نتیجه که می‌نیم‌کننده میانگین تابع زیان درجه دوم خطا، امید ریاضی $E(\theta|y)$ است، یک نتیجه استاندارد در آمار بیزی است. به عنوان مثال، برنارد و اسمیت (۱۹۹۴) را ببینید. برای محاسبه مقدار می‌نیم، داریم

$$\begin{aligned} E((\theta - E(\theta|y))^T A (\theta - E(\theta|y))) &= tr \{ E((\theta - E(\theta|y))^T A (\theta - E(\theta|y))) \} \\ &= tr \{ A E((\theta - E(\theta|y))(\theta - E(\theta|y))^T) \} \\ &= tr(A\Sigma). \end{aligned}$$

برای قسمت ب) کافی است تا نشان دهیم زمانی که $\epsilon \rightarrow 0$ ، $E_{ABC}(\theta|s) = E(\theta|y)$. بنا به تعریف و وقتی $\epsilon \rightarrow 0$ ، با احتمال ۱، $\pi_{ABC}(\theta|s_y) = \pi(\theta|s)$ ، می‌گردد. علاوه بر این

$$E_{ABC}(\theta|s) = \int \theta \pi(\theta|s) d\theta,$$

از طرفی چنانچه $\epsilon \rightarrow 0$ آنگاه

$$\pi(\theta|s) \xrightarrow{P} \pi(\theta|y).$$

بنابراین برای تمام y هایی که با s سازگار هستند (یعنی منجر به تولید s می‌شوند)، رابطه $\int \theta \pi(\theta|s_y) d\theta = E(\theta|y)$ برقرار خواهد بود. بنابراین نتیجه مورد نظر حاصل می‌شود. □

استفاده از تابع زیان توان دوم خطا باعث می‌شود تا میانگین پسین حاصل از تقریب ABC با میانگین پسین واقعی مشابه باشد. استفاده از سایر توابع زیان، باعث تطبیق ویژگی‌های دیگری از

^{۱۲}Risk function

^{۱۰}Positive definite

^{۱۱}Full rank

توزیع‌های پسین ABC و پسین واقعی می‌گردد: برای مثال، تابع زیان قدر مطلق خطا^{۱۳} به تطبیق میانه‌های پسین می‌انجامد. البته در این حالت، ممکن است آماره‌های خلاصه‌شده دیگری میانگین تابع زیان را می‌نیم کنند. البته چنین آماره‌هایی با بعدی بزرگتر از بعد پارامترها، $d > p$ ، باعث بزرگ شدن خطای مونت کارلو می‌شوند (لم ۱۰۲۰۳ را ببینید).

نتیجه ۳۰۲۰۳. در صورت وجود آماره بسنده، از آن به عنوان آماره مورد نظر در تقریب ABC استفاده می‌کنیم. زیرا آماره بسنده تمام اطلاعات موجود در نمونه در مورد پارامتر را دارد. اما به دلیل این که معمولا با مدل‌هایی سر و کار داریم که غالبا در آن‌ها آماره بسنده وجود ندارد، باید از آماره خلاصه‌شده غیربسنده استفاده کرد.

رابرت (۲۰۱۴) در مثالی به تفاوت نتایج عملی استفاده از آماره‌های غیربسنده در مقایسه با آماره‌های بسنده پرداخت. در این جا مشابه مثال رابرت را مورد تحلیل قرار می‌دهیم.

مثال ۴۰۲۰۳. در این مثال ۱۰۰ مشاهده از توزیع $N(\theta_1 = 0, \theta_2 = 1)$ ، به عنوان توزیع واقعی داده‌ها، تولید کردیم. فرض کردیم $\theta_1 \sim N(100, sd = 10)$ و $\theta_2 \sim \text{Gamma}(2, 5)$. در این مثال از میانگین و واریانس به عنوان آماره‌های بسنده و از آماره‌های میانه و mad به عنوان آماره‌های غیربسنده استفاده کردیم. از تقریب RS-ABC با استفاده از 10^6 داده شبیه‌سازی، برای انجام محاسبات کمک گرفتیم. شکل ۱۰۳ نتایج را برای هر دو دسته از آماره‌ها نمایش می‌دهد. در این شکل، بیان دو نکته را لازم می‌دانیم:

۱. با توجه به شکل ۱۰۳، توزیع پسین پارامترها زمانی که از میانگین و واریانس به عنوان آماره‌های خلاصه‌شده استفاده شده است، پهن‌تر است. علت آن هم بسنده بودن میانگین و واریانس است. به این معنی که عدم قطعیت واقعی برآوردگرها بهتر توصیف شده است.

۲. استفاده از آماره‌های بسنده و غیربسنده، در این مثال، تفاوت زیادی در نتیجه استنباط ایجاد نکرده است. بنابراین استفاده از آماره‌های غیربسنده‌ای که بعد کوچکی دارند، می‌تواند نتایج قابل اعتمادی را حاصل دهد. این مساله را به‌طور نظری در مباحث قبلی تشریح کرده‌ایم.

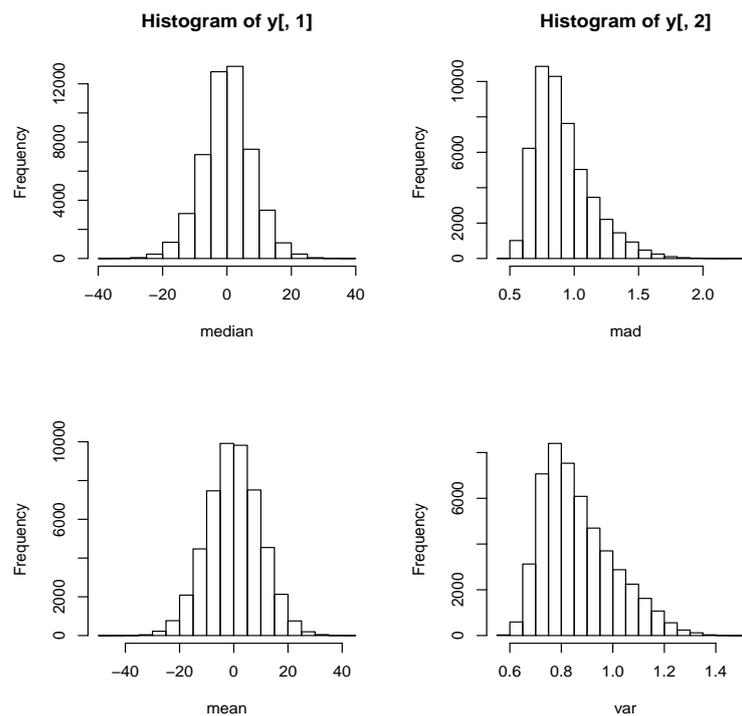
سایر انتخاب‌های آماره خلاصه‌شده

مطالعات مختلفی در مورد انتخاب آماره خلاصه‌شده انجام شده‌اند. به عنوان نمونه، جویس و ماجروم (۲۰۰۸)، اساس انتخاب را بر مفهوم تقریب بسندگی قرار دادند. نونس و بالدینگ (۲۰۱۰) به بررسی مفهوم آنتروپی^{۱۴} و میانگین توان دوم خطا پرداختند. ارزیابی و اجرای هر یک از این روش‌ها نیاز به شبیه‌سازی دارد که در روشی موسوم به ABC نیمه خودکار (فرنهید و پرانگل، ۲۰۱۰) به تشریح به آن پرداخته شده است. بلام (۲۰۱۰b) تاکیدش را برای انتخاب آماره خلاصه‌شده بر روی روش نمونه‌گیری رد و استفاده از روش درست‌نمایی تجربی^{۱۵} قرار داد. اجرای این روش‌ها برای متغیرهای توضیحی در

^{۱۵}Empirical likelihood

^{۱۳}Absolute error loss

^{۱۴}Entropy



شکل ۱.۳: نمودار چگالی پسین حاشیه‌ای پارامترهای (θ_1, θ_2) بر اساس آماره‌های خلاصه‌شده میانه و mad (دو هیستوگرام بالایی) و میانگین و واریانس (دو هیستوگرام پایینی).

مدل‌های رگرسیونی سخت است، زیرا کاوش همه زیرمجموعه‌های ممکن متغیرها کار سخت و زمان‌بری است.

وگمن و همکاران (۲۰۰۹) آماره خلاصه‌شده را با استفاده از طرحی که مشابه ABC نیمه خودکار بود، تولید کردند. در رهیافت آن‌ها مجموعه‌ای از بردارهای پارامتر از توزیع پیشین و داده‌های مرتبط از مدل شبیه‌سازی می‌شوند. سپس با استفاده از روش کمترین توان‌های دوم جزئی^{۱۶} همزمان برآورد پارامترها به عنوان انتخاب آماره‌های خلاصه‌شده و کاهش بعد انجام می‌شوند. در این روش، انتخاب از بین ترکیب‌های مختلف متغیرهای توضیحی، بر اساس معیار اعتبار سنجی متقابل^{۱۷} تعیین می‌گردد. استفاده از رگرسیون خطی موضعی^{۱۸} نیز، در این روش، باعث می‌شود تا همبستگی بین آماره‌ها از بین برود.

۳.۲.۳ خطای تقریب پسین در استفاده از آستانه تحمل (یا هسته) و نرم

همان‌طور که اشاره کردیم، یک خطای غیرقابل اغماض با نمونه‌گیری از $\pi(\theta | \rho(s_x, s_y)) \leq \epsilon$ به جای $\pi(\theta | s_y)$ رخ می‌دهد. در حقیقت با سطح تحمل به اندازه کافی کوچک ϵ و فاصله معقول $\rho(\cdot)$ ، انتظار می‌رود که توزیع $\pi(\theta | \rho(s_x, s_y)) \leq \epsilon$ توزیع واقعی $\pi(\theta | s_y)$ را به خوبی تقریب بزند. اما باید قبول کنیم

^{۱۸}Local linear regression

^{۱۶}Partial least square

^{۱۷}Cross-validation

می‌گردد. از طرفی، به ازای ϵ به اندازه کافی بزرگ، هر نقطه از فضای پارامتر پذیرفته شده و در واقع یک نسخه از توزیع پیشین نتیجه خواهد شد. در زمینه تفاوت $\pi(\theta|\rho(s_x, s_y) \leq \epsilon)$ و $\pi(\theta|s_y)$ مطالعاتی صورت پذیرفته‌اند، که به نتایج نظری برای به دست آوردن کران بالا برای ϵ منتهی شده‌اند. به عنوان نمونه، می‌توان به سیزن و همکاران (۲۰۰۷) و دین و همکاران (۲۰۱۱) اشاره کرد. در ادامه، به اختصار، به تاثیر انتخاب تابع فاصله و سطح تحمل بر تقریب می‌پردازیم.

انتخاب تابع فاصله

توابع فاصله (نرم) مختلف را در فصل اول معرفی کردیم. در متون مربوط به روش‌های ABC، معمولاً از توابع نرم L^p ، به‌ویژه تابع فاصله اقلیدسی، استفاده می‌شود. البته مطالعات تجربی مختلف (پیتر، ۲۰۰۹) نشان داده‌اند که انتخاب تابع فاصله تاثیر قابل ملاحظه‌ای بر روی تقریب حاصل از ABC ندارد.

انتخاب سطح تحمل یا هسته

انتخاب سطح تحمل ϵ ، یک موضوع قدرت محاسباتی است. زیرا ϵ کوچکتر با هزینه محاسبات بیشتر مرتبط است. روش استاندارد انتخاب سطح تحمل (بیومونت و همکاران، ۲۰۰۱) این است که ϵ را به عنوان صدک کوچکی از فاصله شبیه‌سازی شده $\rho(s_y, s_x)$ انتخاب می‌کنند. روش دیگر آن است که بر اساس روش‌های ناپارامتری برآورد تابع چگالی، یک برآورد چگالی برای یک نمونه اولیه از الگوریتم ABC به دست آوریم و سپس ϵ را به عنوان پهنای نوار برآورد مناسب انتخاب کنیم. فیرنهد و پرانگل (۲۰۱۱) دیدگاهشان را مبتنی بر ϵ بهینه که مخالف صفر است، مطرح کردند.

صرف نظر از دیدگاه‌های دیگر در این زمینه، راتمن و همکاران (۲۰۰۹)، روش جدیدی برای خطای تقریب ABC ارائه کردند. آن‌ها ابزار مشخصی در این زمینه برای خوبی برازش در یک مدل معین پیشنهاد کردند. ایده اساسی آن‌ها این بود که سطح تحمل ϵ را به عنوان یک پارامتر اضافی از مدل در توزیع پسین توام وارد کرده و آن را نیز بر مبنای توزیع

$$\pi(\theta, \epsilon|y) \propto \xi(\epsilon|y, \theta)\pi_\theta(\theta)\pi_\epsilon(\epsilon), \quad (21.3)$$

برآورد کنند، به طوری که $\xi(\epsilon|y, \theta)$ نقش تابع درستنمایی مدل را ایفا می‌کند. توابع $\pi_\epsilon(\cdot)$ و $\pi_\theta(\cdot)$ نیز توزیع‌های پیشین متناظر با θ و ϵ هستند. در این دیدگاه، $\xi(\epsilon|y, \theta)$ به عنوان تابع چگالی پیشگوی پیشین^{۱۹} $\rho(s_y, s_x)$ ، به شرط θ و y ، است، به طوری که x از $\pi(x|\theta)$ تولید شده باشد. توجه کنید یک رابطه بین توزیع هدف ویلکینسون (۲۰۰۸) یعنی توزیع (۱۵.۲) با توزیع هدف (۲۱.۳) وجود دارد. این دو توزیع هدف معادل هستند اگر $(y - x)$ را با ϵ جایگزین کنیم.

راتمن و همکاران (۲۰۰۹) الگوریتم ABC دیگری که آن را ABC_μ نامیدند، معرفی کردند. این الگوریتم با استفاده از شبیه‌سازی زنجیر MCMC از توزیع پسین توام (۲۱.۳)، $\xi(\epsilon|y, \theta)$ را با یک تقریب هسته ناپارامتری جایگزین می‌کند. در این حالت، از توزیع پسین حاشیه‌ای ϵ برای انتخاب

^{۱۹}Prior predictive density

مقداری برای آن به‌کار می‌برند، به‌طوری‌که به برآزش خوبی از مدل دست یابند. منظور از برآزش خوب، قرار گرفتن پارامترهای θ در یک فاصله معتبر و معقول می‌باشد. از آن‌جا که تاکید راتمن و همکارنش به استفاده دوباره از داده‌ها است و در دیدگاه بیزی استفاده دوباره از داده‌ها مجاز شمرده نمی‌شود، توزیع هدف (یعنی توزیع پسین) فوق‌را، بر اساس توزیع پیشین ϵ و توزیع شرطی ϵ به شرط مشاهدات و پارامترها، به عنوان یک تابع درست‌نمایی، نسبت به (ϵ, θ) ، تعریف می‌کنند. بنابراین می‌توان از آن، بر پایه الگوریتم‌های ABC، شبیه‌سازی کرد. اما در تفسیر بیزی استنباط‌های نتیجه‌شده باید دقت لازم را خرج کرد؛ زیرا به نظر می‌رسد توزیع پیشین ϵ نقش مهمی در ارزیابی نهایی مدل دارد.

جنبه مهم این بحث، نقش دوگانه خطای تقریب است. به عبارت دیگر، انگیزه اصلی روش‌های ABC، انجام استنباط روی پارامترهای یک مدل است، در حالی‌که، در این جا، هدف از اجرای ABC تعیین مدل است.

اکنون از منظر دیگری به مساله انتخاب سطح تحمل نگاه می‌کنیم و آن را با انتخاب هسته بررسی می‌کنیم. پرانگل (۲۰۱۱) نشان داد، تحت شرایطی معین، در الگوریتم ABC نوفه‌دار، با انتخاب $s_y = E(\theta|y)$ و برای ϵ کوچک، امید ریاضی تابع زیان توان دوم خطا به ازای $\hat{\theta} = E_{ABC}(\theta|s)$ به صورت زیر است:

$$E(L(\theta, \hat{\theta}; A)|y) = \text{trac}(A\Sigma) + \epsilon^2 \int x^T Ax \mathcal{K}(x) dx + o(\epsilon^2). \quad (22.3)$$

همچنین امید ریاضی تابع زیان توان دوم خطا برای ABC استاندارد به صورت زیر است:

$$E(L(\theta, \hat{\theta}; A)) = \text{trac}(A\Sigma) + O(\epsilon^2). \quad (23.3)$$

تعمیم این نتایج ما را به انتخاب هسته مورد نظر رهنمون می‌کند. لم زیر بر اساس (۲۲.۳) نتیجه می‌شود.

لم ۵.۲.۳. (پرانگل، ۲۰۱۱) فرض کنید ماتریس A پرتبه باشد. از بین هسته‌های با میانگین صفر، هسته‌ای که $\int x^T Ax \mathcal{K}(x) dx$ را روی ناحیه $x^T Ax < c$ می‌نیم می‌کند، هسته یکنواخت است، به‌طوری‌که c بر اساس $\int \mathcal{K}(x) dx = 1$ تعیین می‌گردد.

با توجه به لم ۵.۲.۳، پیشنهاد می‌شود برای بعضی از c ها، روی بیضی $x^T Ax < c$ ، از هسته یکنواخت استفاده گردد. پرانگل (۲۰۱۱) برای ABC استاندارد نیز استفاده از یک هسته یکنواخت بر روی یک بیضی را پیشنهاد داد. در عمل محاسبه این بیضی مشکل است. با این حال انتخاب هسته تاثیر کمی بر روی دقت الگوریتم ABC دارد.

۳.۳ نرخ همگرایی

هنگامی که روش ABC در عمل مورد استفاده قرار می‌گیرد، مطالعه جدی روی ویژگی برآوردگرهای ABC فرای انجام روش است. یکی از این ویژگی‌ها نرخ همگرایی است. یعنی برآوردگر حاصل از روش ABC به مقدار دقیق، تحت شرایطی، با چه سرعتی همگرا است. در این بخش، نشان می‌دهیم

اریبی برآوردگرهای مبتنی بر ABC، به طور مجانبی، وقتی $\epsilon \downarrow 0$ ، متناسب است با ϵ^2 . در مقابل، با کاهش یافتن ϵ ، هزینه‌های محاسباتی برای تولید یک نمونه ABC افزایش می‌یابد، به ویژه هنگامی که بعد مشاهدات افزایش یابد. این هزینه محاسباتی متناسب است با ϵ^{-q} ، به طوری که q بعد مشاهدات است. نرخ همگرایی با برقرار کردن تعادلی بهینه بین میانگین توان دوم خطا^{۲۰} (MSE) در مقابل هزینه محاسباتی به دست می‌آید. نتایج حاصل را می‌توان برای انتخاب پارامتر سطح تحمل ϵ نیز مورد استفاده قرار داد.

با توجه به آن چه که تاکنون طرح کردیم، خطای یک برآورد ABC تحت تاثیر: اریبی نمونه‌های ABC است، که با پارامتر سطح تحمل ϵ کنترل می‌گردد، خطای مونت کارلو، که با تعداد نمونه‌های پذیرفته شده کنترل می‌گردد، و خطایی که تحت تاثیر استفاده نکردن از آماره‌های بسنده حاصل می‌شود، که با استفاده از آماره‌های خلاصه شده مناسب به حداقل می‌رسد.

در صورتی که تابع هزینه متناسب با $cost^{-1/2}$ باشد، خطای مونت کارلو کاهش می‌یابد (ووس، ۲۰۱۴). در این جا منظور از $cost$ هزینه است و توان $\frac{1}{2}$ - مستقل از بعد است. در مقابل آن چه در قضیه ۶.۳.۳ خواهد آمد، نشان می‌دهد، تحت انتخاب بهینه ϵ ، در روش ABC، رابطه

$$error \sim cost^{-2/(q+4)}, \quad (24.3)$$

برقرار است.

در ادامه، فرع ۸.۳.۳ نشان می‌دهد این نتیجه، چه تعداد نمونه‌های پذیرفته شده را ثابت نگه داریم، که با کنترل دقت برآورد و تصادفی بودن هزینه حاصل می‌شود، چه تعداد پیشنهادها را ثابت نگه داریم، که با تصادفی بودن نمونه‌های تصادفی میسر می‌گردد، برقرار خواهد ماند. به طور نظری، ثابت نگه داشتن تعداد نمونه‌های پذیرفته شده رضایت بخش و بهتر است. اما در عمل و در اکثر موارد، هزینه محاسباتی، یعنی تعداد پیشنهادها، را ثابت نگه می‌دارند. در هر صورت، نرخ خطا، با افزایش بعد q ، با توجه به (۲۴.۳)، بدتر می‌شود. حتی در حالت یک بعدی، یعنی برای $q = 1$ ، توان $\frac{2}{5}$ - از توان $\frac{1}{2}$ -، برای خطای مونت کارلو، بدتر است. فیرنهد و پرانگل (۲۰۱۲) توان مشابهی را برای آماره خلاصه شده $S(x) = E(\theta|x)$ به دست آوردند. برای برآورد تابع چگالی پسین، بلام (۲۰۱۰) توان کمی بدتر $2/(q+5)$ - را گزارش کرد. تفاوت نتیجه بلام و سایرین در (۲۴.۳) به این دلیل بود که در رهیافت او بر مبنای برآورد تابع چگالی هسته، یک پارامتر پهنای نوار بیشتر مورد بررسی قرار می‌گرفت.

با توجه به این که آماره بسنده بهینه، میانگین توزیع پسین است، در این جا هدف را برآورد امید ریاضی پسین تابعی مانند $h(\theta)$ ، یعنی

$$Y = E(h(\theta)|X = x), \quad (25.3)$$

قرار می‌دهیم، که برآورد آن به صورت

$$Y_n^\epsilon = 1/n \sum_{j=1}^n h(\theta_j^\epsilon), \quad (26.3)$$

^{۲۰} Mean squared error

است، به طوری که θ_j ها پارامترهای پذیرفته شده از الگوریتم‌های مختلف ABC است و اندیس بالای ϵ برای تاکید سطح تحمل در روش ABC است. وقتی $n \rightarrow \infty$ ، برآورد Y_n^ϵ دقیقاً به مقدر واقعی Y همگرا نمی‌شود. کیفیت تقریب را می‌توان با کاهش ϵ بهبود بخشید، اما این کاهش منجر به احتمال پذیرش بیشتر در الگوریتم رد و در نتیجه افزایش هزینه محاسباتی برای برآورد Y_n^ϵ می‌گردد. پس یک داد و ستد بین دقت در نتایج و هزینه محاسبات وجود دارد. در ادامه، دو قضیه مهم در این زمینه را بیان خواهیم کرد.

همان‌طور که در بالا اشاره کردیم، دو دیدگاه متفاوت برای اندازه نمونه استفاده شده در محاسبه برآورد Y_n^ϵ ABC وجود دارد. اگر N تعداد پیشنهاد‌های مورد نیاز برای n نمونه خروجی، با $N \geq n$ باشد، یک روش آن است که تعداد N پیشنهاد را ثابت در نظر بگیریم. در این حالت، تعداد n نمونه پیشنهادی پذیرفته شده، تصادفی خواهد بود. در مقابل، برای n ثابت مشخص، با ایجاد حلقه در الگوریتم ABC می‌توان n نمونه پذیرفته شده را با یک N تصادفی به دست آورد. در ادامه، دیدگاه N ثابت را دنبال می‌کنیم.

۱.۳.۳ نتایج نظری نرخ همگرایی

با توجه به خلاصه‌ای که از نرخ همگرایی، در بخش کاربردی روش ABC، گفتیم، نیازی به دانستن توزیع‌های θ ، x و S نیست. زیرا می‌توان از این متغیرها به تعداد دلخواه نمونه تولید کرد. فرض کنید تابع چگالی توام θ و S وجود دارد و به صورت $f_{S,\theta}(\cdot, \cdot)$ است. همچنین توابع چگالی حاشیه‌ای S ، θ و چگالی شرطی $\theta|S$ وجود دارند و به صورت زیر می‌باشند:

$$\begin{aligned} f_S(s) &= \int_{\mathbb{R}^p} f_{S,\theta}(s,t) dt & s \in \mathbb{R}^q \\ f_\theta(t) &= \int_{\mathbb{R}^q} f_{S,\theta}(s,t) ds & t \in \mathbb{R}^p \\ f_{\theta|S}(t|s) &= \begin{cases} \frac{f_{S,\theta}(s,t)}{f_S(s)} & \text{اگر } f_S(s) > 0 \\ 0 & \text{در غیر این صورت} \end{cases} \end{aligned}$$

هدف همگرایی Y_n^ϵ به Y وقتی $n \rightarrow \infty$ است.

تعریف ۱.۳.۳. فرض کنید $h: \mathbb{R}^p \rightarrow \mathbb{R}$ تابعی باشد که $E(|h(\theta)|) < \infty$. برای هر $s \in \mathbb{R}^q$ تابع $\varphi_h: \mathbb{R}^q \rightarrow \mathbb{R}$ را به صورت

$$\varphi_h(s) = \int_{\mathbb{R}^p} h(t) f_{S,\theta}(s,t) dt,$$

و برای هر $s^* \in \mathbb{R}^q$ تابع $\varphi_h^{(\epsilon)}: \mathbb{R}^q \rightarrow \mathbb{R}$ را به صورت

$$\varphi_h^{(\epsilon)}(s^*) = \frac{1}{|B(s^*, \epsilon)|} \int_{B(s^*, \epsilon)} \varphi_h(s) ds,$$

تعریف می‌کنیم، که در آن $|B(s^*, \epsilon)|$ نشان‌دهنده حجم کره $B(s^*, \epsilon)$ ، به مرکز s^* و شعاع ϵ است.

با استفاده از تعریف ۱.۳.۳ برای $h(\theta) = 1$ ، $\varphi_1(s) = f_S(s)$ ، در حالت کلی، برای تابع دلخواه $h(\cdot)$ ، خواهیم داشت $\varphi_h(s) = f_S(s)E(h(\theta)|S = s)$. مقدار دقیق Y در (۲۵.۳) و میانگین برآوردگر Y_n^ϵ را می‌توان با تعریف ۱.۳.۳ بازگو کرد. این بازنویسی را در زیر به صورت یک لم بیان می‌کنیم.

لم ۲.۳.۳. فرض کنید $h: \mathbb{R}^p \rightarrow \mathbb{R}$ باشد به طوری که $E(|h(\theta)|) < \infty$. برای f_S و تقریباً هر $s^* \in \mathbb{R}^q$

$$E(h(\theta)|S = s^*) = \frac{\varphi_h(s^*)}{\varphi_1(s^*)}$$

و

$$E(Y_n^\epsilon) = E(h(\theta_1^\epsilon)) = \frac{\varphi_h^{(\epsilon)}(s^*)}{\varphi_1^{(\epsilon)}(s^*)}.$$

برهان. با فرض $E(|h(\theta)|) < \infty$ می‌توان نتیجه گرفت

$$\begin{aligned} \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} |h(t)| f_{S,\theta}(s, t) dt ds &= \int_{\mathbb{R}^p} |h(t)| \int_{\mathbb{R}^q} f_{S,\theta}(s, t) ds dt \\ &= \int_{\mathbb{R}^p} |h(t)| f_\theta(t) dt \\ &= E(|h(\theta)|) < \infty. \end{aligned}$$

بنابراین با توجه به این‌که برای تقریباً هر $s \in \mathbb{R}^q$ ، $\int_{\mathbb{R}^p} |h(t)| f_{S,\theta}(s, t) dt < \infty$ ، در نتیجه میانگین شرطی $E(h(\theta)|S = s^*)$ ، برای هر f_S و تقریباً هر $s^* \in \mathbb{R}^q$ وجود دارد. با استفاده از قضیه بیز، داریم

$$\begin{aligned} E(h(\theta)|S = s^*) &= \int_{\mathbb{R}^p} h(t) f_{\theta|s}(t|s^*) dt \\ &= \int_{\mathbb{R}^p} h(t) \frac{f_{S,\theta}(s^*, t)}{f_S(s^*)} dt \\ &= \frac{1}{f_S(s^*)} \int_{\mathbb{R}^p} h(t) f_{S,\theta}(s^*, t) dt \\ &= \frac{\varphi_h(s^*)}{\varphi_1(s^*)}. \end{aligned}$$

به عبارت دیگر، نمونه‌های θ_j^ϵ مطابق با توزیع شرطی θ به شرط $S \in B(s^*, \epsilon)$ ، توزیع شده‌اند. بنابراین تابع چگالی نمونه‌ای θ_j^ϵ را می‌توان به صورت زیر نوشت:

$$f_{\theta_j^\epsilon}(t) = \frac{1}{Z} \frac{1}{|B(s^*, \epsilon)|} \int_{B(s^*, \epsilon)} f_{S,\theta}(s, t) ds$$

که در آن، ثابت نرمال‌ساز Z به صورت زیر تعریف می‌شود:

$$\begin{aligned} Z &= \frac{1}{|B(s^*, \epsilon)|} \int_{B(s^*, \epsilon)} f_S(s) ds = \frac{1}{|B(s^*, \epsilon)|} \int_{B(s^*, \epsilon)} \varphi_1(s) ds \\ &= \varphi_1^{(\epsilon)}(s^*). \end{aligned}$$

بنابراین

$$\begin{aligned}
 E(Y_n^\epsilon) &= E(h(\theta^\epsilon)) \\
 &= \int_{\mathbb{R}^p} h(t) f_{\theta^\epsilon}(t) dt \\
 &= \int_{\mathbb{R}^p} \frac{1}{Z} \frac{1}{|B(s^*, \epsilon)|} \int_{B(s^*, \epsilon)} f_{S, \theta}(s, t) h(t) ds dt \\
 &= \frac{1}{\varphi_1^{(\epsilon)}(s)} \int_{B(s^*, \epsilon)} \frac{1}{|B(s^*, \epsilon)|} \underbrace{\int_{\mathbb{R}^p} h(t) f_{S, \theta}(s, t) dt}_{\varphi_h(s)} ds \\
 &= \frac{1}{\varphi_1^{(\epsilon)}(s)} \underbrace{\frac{1}{|B(s^*, \epsilon)|} \int_{B(s^*, \epsilon)} \varphi_h(s) ds}_{\varphi_h^{(\epsilon)}(s^*)} \\
 &= \frac{\varphi_h^{(\epsilon)}(s^*)}{\varphi_1^{(\epsilon)}(s^*)}.
 \end{aligned}$$

□

قضیه ۳.۳.۳. (باربر و همکاران، ۲۰۱۴) فرض کنید $h : \mathbb{R}^p \rightarrow \mathbb{R}$ باشد، به طوری که $E(|h(\theta)|) < \infty$ برای هر f_S و تقریباً هر $s^* \in \mathbb{R}^q$ ، برآورد ABC Y_n^ϵ با عبارت (۲۶.۳)، در دو شرط زیر صدق می‌کند:

$$1. \text{ برای هر } \epsilon > 0 \quad Y_n^\epsilon \xrightarrow{a.s.} E(Y_n^\epsilon)$$

$$2. \text{ برای هر } n \in \mathcal{N} \quad \lim_{\epsilon \downarrow 0} E(Y_n^\epsilon) = E(h(\theta) | S = s^*)$$

برهان. برای اثبات قسمت اول، چون $E(|h(\theta)|) < \infty$ ، در نتیجه

$$E(|Y_n^\epsilon|) \leq E(|h(\theta)|) = \frac{\varphi_{|h|}(s^*)}{\varphi_1(s^*)} < \infty.$$

زمانی که $\varphi_1(s^*) = f_S(s^*)$ بنا بر قانون قوی اعداد بزرگ، Y_n^ϵ به صورت قریب به یقین (*a.s.*) به $E(Y_n^\epsilon)$ همگرا می‌شود.

برای حالت دوم، چون $\varphi_1 = f_S \in L^1(\mathbb{R}^q)$ ، می‌توان با استفاده از قضیه مشتق‌پذیری لبگ^{۲۱} (رودین، ۱۹۸۷) نتیجه گرفت، هنگامی که $\epsilon \downarrow 0$ ، برای هر $s^* \in \mathbb{R}^q$ ، $\varphi_1^{(\epsilon)}(s^*) \rightarrow \varphi_1(s^*)$ به طور مشابه، چون

$$\begin{aligned}
 \int_{\mathbb{R}^q} |\varphi_h(s)| ds &\leq \int_{\mathbb{R}^p} |h(t)| \int_{\mathbb{R}^q} f_{S, \theta}(s, t) ds dt \\
 &= \int_{\mathbb{R}^p} |h(t)| f_\theta(t) dt < \infty,
 \end{aligned}$$

^{۲۱}Lebesgue differentiation theorem

بنابراین $\varphi_h \in L^1(\mathbb{R}^q)$ ، و وقتی $\epsilon \downarrow 0$ برای تقریباً هر $s^* \in \mathbb{R}^q$ ، $\varphi_h^{(\epsilon)}(s^*) \rightarrow \varphi_h(s^*)$ ، با استفاده از لم ۲.۳.۳ و تقریباً هر $s^* \in \mathbb{R}^q$ داریم

$$\begin{aligned} \lim_{\epsilon \downarrow 0} E(Y_n^\epsilon) &= \frac{\varphi_h^{(\epsilon)}(s^*)}{\varphi_1^{(\epsilon)}(s^*)} \\ &= \frac{\varphi_h(s^*)}{\varphi_1(s^*)} \\ &= E(h(\theta)|S = s^*). \end{aligned}$$

□

قضیه ۳.۳.۳ بیان‌کننده همگرایی Y_n^ϵ به Y است، اما در مورد نرخ این همگرایی حرفی نمی‌زند. قضیه مهم ۶.۳.۳، در ادامه، سرعت این همگرایی را مشخص می‌کند. بر این اساس، در مورد MSE برآوردگر Y_n^ϵ و رابطه آن با هزینه محاسبه برآورد Y_n^ϵ ، بحث خواهیم کرد. معیار MSE برآوردگر مورد نظر عبارتست از

$$MSE(Y_n^\epsilon) = E\left((Y_n^\epsilon - Y)^2\right) = Var(Y_n^\epsilon) + bias(Y_n^\epsilon)^2. \quad (27.3)$$

زمان (هزینه) لازم برای برآورد Y_n^ϵ را، در الگوریتم ABC، با $Cost(Y_n^\epsilon)$ نشان می‌دهیم. این هزینه متناسب است با $a + bN$ ، که در آن a و b ثابت هستند و N تعداد تکرارهای الگوریتم ABC مورد نظر تا برقرار شدن شرط $j = n$ است. یا به عبارتی، N تعداد پیشنهادهای مورد نیاز تا تولید n نمونه پذیرفته شده می‌باشد. برای شرح رفتار MSE و Cost در قضیه مهم ۶.۳.۳ از تعریف ۴.۳.۳ و فرض ۵.۳.۳ استفاده می‌کنیم.

تعریف ۴.۳.۳ برای دنباله‌های $(a_n)_{n \in \mathbb{N}}$ و $(b_n)_{n \in \mathbb{N}}$ از اعداد حقیقی مثبت، $a_n \sim b_n$ ، هرگاه ثابت c ، که $0 < |c| < \infty$ ، وجود داشته باشد، به طوری که

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = c.$$

دقت کنید این تعریف، معادل است با تعریف مرتبه یک دنباله که با نماد $O(\cdot)$ نمایش داده می‌شود.

فرض ۵.۳.۳ چگالی $f_S(\cdot)$ و تابع $H(s) = \int_{\mathbb{R}^p} h(t) f_{S,\theta}(s,t) dt$ ، در یک همسایگی از $s^* \in \mathbb{R}^q$ ، دارای مشتق مرتبه سوم پیوسته هستند.

قضیه ۶.۳.۳ (باربر و همکاران، ۲۰۱۴) فرض کنید برای $h: \mathbb{R}^p \rightarrow \mathbb{R}$ ، $E(h(\theta)^2) < \infty$ و فرض کنید $S(\cdot)$ بسنده باشد و در فرض ۵.۳.۳ صدق کند. اگر $Var(h(\theta)|S = s^*) > 0$ و

$$C(s^*) = \frac{\Delta\varphi_h(s^*) - Y \cdot \Delta\varphi_1(s^*)}{2(q+2)\varphi(s^*)} \neq 0, \quad (28.3)$$

که در آن عبارت $\Delta\varphi_h(\cdot)$ نشان‌دهنده عملگر لاپلاس برای تابع $\varphi_h(\cdot)$ است. آن‌گاه برای f_S به ازای هر s^* ، دو بند زیر برقرار هستند:

۱. اگر $(\epsilon_n)_{n \in \mathcal{N}}$ یک دنباله باشد، به طوری که $\epsilon_n \sim n^{-1/4}$ ، چنانچه $n \rightarrow \infty$ ،
 $MSE(Y_n^\epsilon) \sim E(Cost(Y_n^\epsilon))^{-\frac{4}{q+4}}$.

۲. توان $-\frac{4}{(q+4)}$ در قسمت قبل بهینه شده است. یعنی برای دنباله $(\epsilon_n)_{n \in \mathcal{N}}$ با $\epsilon_n \downarrow 0$ ، وقتی
 $n \rightarrow \infty$

$$\liminf_{n \rightarrow \infty} \frac{MSE(Y_n^{\epsilon_n})}{E(Cost(Y_n^{\epsilon_n}))^{-\frac{4}{q+4}}} > 0.$$

برهان. اثبات قضیه فوق را در قالب سه لم، ۹.۳.۳، ۱۰.۳.۳ و ۱۱.۳.۳ بیان می‌کنیم. □

برای اثبات قضیه ۶.۳.۳ ابتدا اریبی برآوردگر Y_n^ϵ را بررسی می‌کنیم. این اریبی به اندازه نمونه وابسته نیست. وابستگی به سطح تحمل ϵ در لم ۷.۳.۳ مطرح شده است. این لم بخش کلیدی برهان قضیه ۶.۳.۳ است و اثبات آن را به استورات و همکاران (۲۰۱۴) ارجاع می‌دهیم.

لم ۷.۳.۳ فرض کنید $E(|h(\theta)|) < \infty$ و S در فرض ۵.۳.۳ صدق کند. آنگاه برای f_S به ازای تقریباً هر $s^* \in \mathbb{R}^q$ داریم

$$bias(Y_n^\epsilon) = C(s^*)\epsilon^2 + O(\epsilon^3),$$

که در آن $\epsilon \downarrow 0$ و ثابت $C(s^*)$ در (۲۸.۳) داده شده است.

نتیجه ۸.۳.۳. فرض کنید $E(h(\theta)^2) < \infty$. آنگاه برای f_S به ازای تقریباً هر $s^* \in \mathbb{R}^q$ و به طور یکنواخت در n ، داریم

$$\lim_{\epsilon \downarrow 0} nVar(Y_n^\epsilon) = Var(h(\theta)|S = s^*).$$

برهان. از تعریف واریانس، داریم

$$\begin{aligned} Var(Y_n^\epsilon) &= \frac{1}{n} Var(h(\theta_j^\epsilon)) \\ &= \frac{1}{n} \left(E(h(\theta_j^\epsilon)^2) - E(h(\theta_j^\epsilon))^2 \right). \end{aligned}$$

با فرض قضیه ۳.۳.۳ و با استفاده از $h^2(\cdot)$ ، $E(h(\theta)^2|S = s^*)$ ، $\lim_{\epsilon \downarrow 0} E(h(\theta_j^\epsilon)^2) = E(h(\theta)^2|S = s^*)$ ، از آنجایی که $E(h(\theta)^2) < \infty$ ، پس $E(|h(\theta)|) < \infty$ همچنین

$$\lim_{\epsilon \downarrow 0} E(h(\theta_j^\epsilon)) = E(h(\theta)|S = s^*).$$

بنابراین

$$\begin{aligned} \lim_{\epsilon \downarrow 0} nVar(Y_n^\epsilon) &= E(h(\theta)^2|S = s^*) - E(h(\theta)|S = s^*)^2 \\ &= Var(h(\theta)|S = s^*). \end{aligned}$$

□

برای اثبات قسمتی از قضیه ۶.۳.۳، اریبی Y_n^ϵ با هزینه محاسباتی برای محاسبه Y_n^ϵ برابر شده است. هنگامی که ϵ کاهش می‌یابد، نمونه کمتری با شرط $\rho(\cdot) < \epsilon$ پذیرفته می‌شوند. در نتیجه، زمان اجرای الگوریتم افزایش خواهد یافت. لم ۹.۳.۳ این حالت را بررسی می‌کند.

لم ۹.۳.۳. فرض کنید f_S در s^* پیوسته باشد. متوسط هزینه محاسباتی برای محاسبه برآورد Y_n^ϵ برای $n \in \mathcal{N}$ و $\epsilon > 0$ ، به صورت زیر است:

$$E(\text{Cost}(Y_n^\epsilon)) = c_1 + c_2 n \epsilon^{-q} (1 + c_3(\epsilon)),$$

که در آن c_1 و c_2 ثابت هستند و c_3 به n وابسته نیست. همچنین اگر $\epsilon \rightarrow 0$ ، آن‌گاه $c_3(\epsilon) \rightarrow 0$.

برهان. هزینه محاسباتی برای اجرای الگوریتم ABC به صورت $a + bN$ است که a و b ثابت هستند، و N تعداد تکرارهای تصادفی لازم تا پذیرش n نمونه است. تعداد تکرارهای مورد نیاز در یک نمونه ABC دارای توزیع هندسی با پارامتر $(B(s^*, \epsilon))$ است. بنابراین N که مجموع n مقدار مستقل است، دارای توزیع هندسی با میانگین $E(N) = n/p$ است. احتمال p را می‌توان به صورت زیر نوشت:

$$\begin{aligned} p &= \int_{B(s^*, \epsilon)} f_S(s) ds \\ &= |B(s^*, \epsilon)| \varphi_1^{(\epsilon)}(s^*) \\ &= \epsilon^q |B(s^*, 1)| \cdot \varphi_1^{(\epsilon)}(s^*). \end{aligned}$$

از آنجایی که $\varphi_1 = f_S$ در s^* پیوسته است، پس اگر $\epsilon \rightarrow 0$ ، آن‌گاه $\varphi_1^{(\epsilon)}(s^*) \rightarrow \varphi_1(s^*)$. بنابراین برای مقدار مثبت c ، $p = c\epsilon^q(1 + o(1))$. در نتیجه، هزینه محاسباتی به صورت زیر تبدیل می‌شود:

$$\begin{aligned} E(\text{Cost}(Y_n^\epsilon)) &= a + bE(N) \\ &= a + b \frac{n}{c\epsilon^q(1 + o(1))} \\ &= a + \frac{b}{c} n \epsilon^{-q} \frac{1}{(1 + o(1))}. \end{aligned}$$

□

در انتها، دو لم زیر تقریب خطا را از یک طرف به اریبی و از طرف دیگر به متوسط هزینه محاسباتی، پیوند می‌دهند.

لم ۱۰.۳.۳. فرض کنید، برای $n \in \mathcal{N}$ ، $\epsilon_n \sim n^{-1/4}$. همچنین فرض کنید $S, E(h(\theta)^2) < \infty$. فرض ۵.۳.۳ صدق کند و $\text{Var}(h(\theta)|S = s^*)$ کراندار باشد. در این صورت، برای f_S به ازای تقریباً هر $s^* \in \mathbb{R}^q$ ، خطا در $MSE(y_n^\epsilon) \sim n^{-1}$ صدق می‌کند. متوسط هزینه محاسباتی با رابطه $E(\text{Cost}(y_n^\epsilon)) \sim n^{1+q/4}$ افزایش می‌یابد و اگر $n \rightarrow \infty$ ، آن‌گاه $MSE(y_n^\epsilon) \sim E(\text{Cost}(y_n^\epsilon))^{-\frac{4}{q+4}}$.

برهان. با فرض وجود $|D| < \infty$ ، به طوری که $\lim \epsilon_n n^{1/q} = D$ و با استفاده از لم ۷.۳.۳ و فرع ۸.۳.۳، داریم

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{MSE(y_n^\epsilon)}{n^{-1}} &= \lim_{n \rightarrow \infty} n \left(\text{Var}(y_n^\epsilon) + (\text{bias}(y_n^\epsilon))^2 \right) \\ &= \lim_{n \rightarrow \infty} n \text{Var}(y_n^\epsilon) + \lim_{n \rightarrow \infty} n \left(C(s^*) \epsilon_n^2 + O(\epsilon_n^3) \right)^2 \\ &= \text{Var}(h(\theta)|S = s^*) + \lim_{n \rightarrow \infty} \left(C(s^*) + \frac{O(\epsilon_n^3)}{\epsilon_n^2} \right)^2 \cdot (\epsilon_n n^{1/q})^4 \\ &= \text{Var}(h(\theta)|S = s^*) + C(s^*)^2 D^4. \end{aligned}$$

از طرف دیگر، با استفاده از لم ۹.۳.۳ می‌توان نوشت

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{E(\text{Cost}(y_n^\epsilon))}{n^{1+q/4}} &= \lim_{n \rightarrow \infty} \frac{c_1 + c_2 n \epsilon_n^{-q} (1 + c_3 (\epsilon_n))}{n^{1+q/4}} \\ &= 0 + c_2 \lim_{n \rightarrow \infty} \frac{1}{(\epsilon_n n^{1/4})^q} \left(1 + \lim_{n \rightarrow \infty} (1 + c_3 (\epsilon_n)) \right) \\ &= \frac{c_2}{D^q}. \end{aligned}$$

در نهایت با ترکیب نرخ هزینه و خطا داریم:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{MSE(y_n^\epsilon)}{E(\text{Cost}(y_n^\epsilon))^{-\frac{4}{q+4}}} &= \frac{\lim_{n \rightarrow \infty} \frac{MSE(y_n^\epsilon)}{n^{-1}}}{\lim_{n \rightarrow \infty} \left(\frac{E(\text{Cost}(y_n^\epsilon))}{n^{1+q/4}} \right)^{-\frac{4}{q+4}}} \quad (29.3) \\ &= (\text{Var}(h(\theta)|S = s^*) + C(s^*)^2 D^4) \cdot \left(\frac{D^q}{c_2} \right)^{-\frac{4}{q+4}}. \end{aligned}$$

چون طرف راست عبارت فوق غیرصفر و کراندار است، نتیجه حاصل می‌شود. \square

لم ۱۰.۳.۳ بیان‌کننده نرخ بهینه برای کاهش ϵ_n به عنوان یک تابع از n است. با بررسی برهان این لم، می‌توان ثابت بهینه متناظر با آن را به دست آورد: ϵ_n باید به شکل $\epsilon_n = D n^{-1/4}$ باشد، به طوری که $MSE D$ را می‌نیم کند. با این معیار، مقدار بهینه D به صورت زیر به دست می‌آید:

$$D_{opt} = \left(\frac{q \text{Var}(h(\theta)|S = s^*)}{\epsilon C(s^*)^2} \right)^{\frac{1}{4}}, \quad (30.3)$$

واریانس $\text{Var}(h(\theta)|S = s^*)$ به آسانی برآورد می‌گردد، اما به نظر می‌رسد که برآورد $C(s^*)$ در ABC کار راحتی نباشد. لم ۱۱.۳.۳ اثبات قضیه ۶.۳.۳ را کامل می‌کند و نشان می‌دهد سایر انتخاب‌های ϵ_n ، برای تابع هزینه مشابه، نمی‌تواند منجر به نرخ همگرایی بهتری برای خطای تقریب گردد.

لم ۱۱.۳.۳. برای $n \in \mathcal{N}$ ، فرض کنید $\epsilon_n \rightarrow 0$. همچنین فرض کنید $E(h(\theta)^2) < \infty$ ، $\text{Var}(h(\theta)|S = s^*)$ کراندار باشد، در فرض ۵.۳.۳ صدق کند و $C(s^*) \neq 0$ باشد. در این صورت، برای f_s به ازای تقریباً هر $s^* \in \mathbb{R}^q$ ، داریم

$$\liminf_{n \rightarrow \infty} \frac{MSE(Y_n^{\epsilon_n})}{E(\text{Cost}(Y_n^{\epsilon_n}))^{-\frac{4}{q+4}}} > 0.$$

برهان. با استفاده از لم ۲.۳.۳ و برای n به اندازه کافی بزرگ، چون $C(s^*)^2 > 0$ پس

$$\begin{aligned} MSE(Y_n^{\epsilon_n}) &= Var(Y_n^{\epsilon_n}) + (bias(Y_n^{\epsilon_n}))^2 \\ &= \frac{Var((\theta_j^{\epsilon_n}))}{n} + (C(s^*)\epsilon_n^2 + O(\epsilon_n^3))^2 \\ &= \frac{Var((\theta_j^{\epsilon_n}))}{n} + C(s^*)^2\epsilon_n^4 + O(\epsilon_n^5) \\ &\geq \frac{Var((\theta_j^{\epsilon_n}))}{n} + \frac{C(s^*)^2}{2}\epsilon_n^4. \end{aligned} \quad (31.3)$$

با استفاده از لم ۹.۳.۳، زمانی که $n \rightarrow \infty$ می‌توان نوشت

$$\begin{aligned} E(Cost(Y_n^{\epsilon_n}))^{-\frac{4}{q+4}} &\sim (n\epsilon_n^{-q})^{-\frac{4}{q+4}} \\ &= (n^{-1})^{\frac{4}{q+4}} (\epsilon_n^4)^{\frac{q}{q+4}} \\ &\sim \left(\frac{4}{q+4} \frac{Var((\theta_j^{\epsilon_n}))}{n} \right)^{\frac{4}{q+4}} \left(\frac{q}{q+4} \frac{C(s^*)^2}{2} \epsilon_n^4 \right)^{\frac{q}{q+4}}, \end{aligned} \quad (32.3)$$

که می‌توان عامل ثابت را، به این دلیل که رابطه \sim شامل مقدار ثابت نیست، در عبارت آخر قرار داد. با استفاده از نامساوی یونگ^{۲۲} داریم

$$\begin{aligned} \frac{Var((\theta_j^{\epsilon_n}))}{n} + \left(\frac{C(s^*)^2}{2} \epsilon_n^4 \right) \\ \geq \left(\frac{4}{q+4} \frac{Var((\theta_j^{\epsilon_n}))}{n} \right)^{\frac{4}{q+4}} \left(\frac{q}{q+4} \frac{C(s^*)^2}{2} \epsilon_n^4 \right)^{\frac{q}{q+4}} \end{aligned} \quad (33.3)$$

با ترکیب سه عبارت (۳۱.۳) و (۳۲.۳) و (۳۳.۳) و برای بعضی از $c > 0$ و هر n به اندازه کافی بزرگ، نتیجه می‌شود

$$MSE(Y_n^{\epsilon_n}) \geq c.E(Cost(Y_n^{\epsilon_n}))^{-\frac{4}{q+4}}.$$

□

با توجه به لم‌های ۹.۳.۳، ۱۰.۳.۳ و ۱۱.۳.۳، قضیه ۶.۳.۳ اثبات می‌شود. سرعت همگرایی در قسمت اول قضیه ۶.۳.۳ باید با سرعت متناظر آن در خطای مونت کارلو مقایسه شود. از آنجایی که برآوردهای مونت کارلو ناریب هستند، زمانی که تعداد نمونه تولیدی برابر n باشد، جذر MSE (RMSE) برای برآورد مونت کارلو متناسب با $1/\sqrt{n}$ است. بنابراین برای برآورد مونت کارلو $RMSE \sim Cost^{-1/2}$ می‌شود. توان متناظر از قضیه ۶.۳.۳ با گرفتن ریشه دوم، که $-(q+4)/2$ است، به دست آمده و این نشان‌دهنده کند بودن همگرایی برآورد ABC است. این کارایی کاهش یافته، در حقیقت،

^{۲۲}Young inequality

نتیجه‌ای از خطای اضافی تولیدشده به وسیله اریبی برآورد ABC است. همچنین هنگامی که هزینه محاسباتی افزایش می‌یابد، ϵ باید به نسبت تا $n^{-1/4}$ کاهش یابد. در این حالت خطا کاهش یافته و متناسب است با هزینه یا با توان $(q+4)/-4$.

دومین قسمت قضیه ۶.۳.۳ نشان می‌دهد هیچ انتخابی از ϵ نمی‌تواند منجر به بهتر شدن (یا منفی‌تر شدن) گردد. اگر سطح تحمل ϵ_n کاهش یابد، در این صورت منجر به بزرگتر شدن توان می‌گردد، که در حد برای مقادیر بزرگ n ، همیشه بدترین انتخاب، $\epsilon_n = cn^{-1/4}$ ، برای $c > 0$ است. این نتیجه فقط برای حالت $n \rightarrow \infty$ به کار می‌رود.

حاصل قضیه ۶.۳.۳ این است که برای یک مساله مفروض، سطح تحمل ϵ کاهش یابد. بهتر است تا برای مقایسه برآورد ABC از q مختلف استفاده گردد. در لم ۲.۳.۳ نشان دادیم که اریبی برآورد ABC متناسب با $\epsilon^2 \sim bias$ است. همچنین در لم ۷.۳.۳ نشان دادیم که در برآورد ABC، $Cost \sim n\epsilon^{-q}$ است. به تناسب این دو لم، در قضیه ۶.۳.۳، اساس انتخاب ϵ بهینه در عبارت (۲۴.۳) است. این یعنی معیاری برای انتخاب سرعت همگرایی با استفاده از ϵ .

فصل ۴

کاربردها

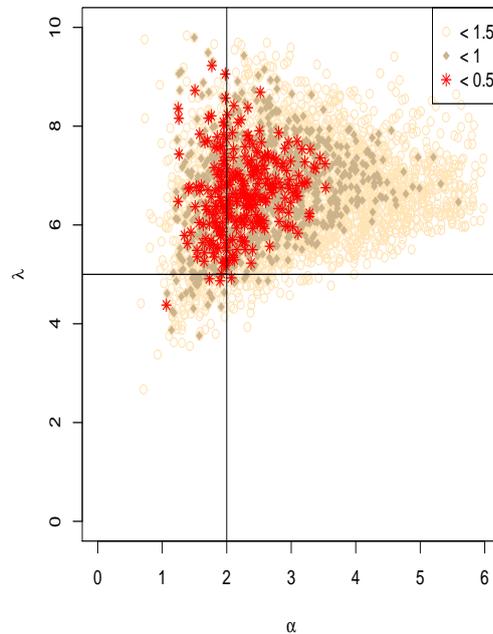
۱.۴ مقدمه

با توجه به بررسی پارامترهای مدل (s, ρ, ϵ) در فصل ۳، دیدیم که کاهش ϵ ما را به توزیع پسین واقعی نزدیک‌تر می‌کند. اما خود این با افزایش هزینه همراه است، و با افزایش بعد، این افزایش به صورت نمایی است. در این فصل به بررسی عملکرد ABC می‌پردازیم و در چند مثال، کاربرد آن را نشان می‌دهیم. در بخش ۱.۱.۴، به توصیف ABC با استفاده از روش $ABC-RS$ می‌پردازیم (این مثال بیشتر برای روشن شدن مطلب می‌باشد). در بخش ۲.۱.۴ مدل $MA(q)$ سری‌زمانی را معرفی می‌کنیم. و نشان می‌دهیم که با توجه به پیچیده بودن شکل تابع درستنمایی از روش $ABC-RS$ استفاده می‌کنیم. با توجه به نرخ رد بالا در $ABC-RS$ از $ABC-MCMC$ استفاده می‌کنیم. خواهیم دید که با کاهش ϵ ، به توزیع پسین واقعی نزدیک می‌شویم. هرچند با این کار هزینه افزایش می‌ابد. در بخش ۳.۱.۴ به بررسی یک مثال کاربردی می‌پردازیم. مدل لوتکا-ولترا را معرفی می‌کنیم. همچنین نشان می‌دهیم که محاسبه تابع درستنمایی آن غیر ممکن است. با استفاده از ABC و شبیه‌سازی نشان می‌دهیم که توزیع پسین به دست آمده به‌ازای کاهش سطح تحمل به پسین مدل واقعی نزدیک می‌شود. از تحلیل روش $Noisy ABC$ استفاده نشده است. آن مشابه با ۳ الگوریتم اصلی، اما با کمی دقت بیشتر است. به ۳ مثال مذکور می‌پردازیم.

۱.۱.۴ توزیع وایبل

برای نشان دادن عملکرد روش ABC ، مدل تصادفی توزیع وایبل را در نظر می‌گیریم. می‌دانیم که تابع درستنمایی توزیع و توزیع پسین آن قابل محاسبه است. با این حال برای راحتی کار از روش $ABC-RS$ استفاده می‌کنیم. ۲۰ داده مشاهده شده از توزیع وایبل به تصادف تولید می‌کنیم و آن را به عنوان مشاهده واقعی در نظر می‌گیریم. پارامترهای توزیع را $\alpha = 2$ و $\lambda = 5$ فرض می‌کنیم. همچنین از آماره خلاصه شده میانگین و انحراف معیار استفاده می‌کنیم. حال بر اساس این مشاهدات و آماره خلاصه شده، داده‌های شبیه‌سازی را تولید می‌کنیم. در ضمن فرض می‌کنیم توزیع پیشین، توزیع یکنواخت است.

با توجه به الگوریتم نمونه گیری رد، نمونه پذیرفته شده را بر اساس چندک‌های $(1/5, 1, 4/5)$ درصد از فاصله مورد نظر، به دست می‌آوریم. نتایج حاصل در شکل ۱.۴ آمده است.



شکل ۱.۴: پارامترهای پذیرفته شده بر اساس چندک‌های $(1/5, 1, 4/5) = \epsilon$ درصد از فاصله مورد نظر، و آماره خلاصه شده میانگین و انحراف معیار

با توجه به شکل ۱.۴، ملاحظه می‌گردد که غالب تمرکز در ناحیه پارامترهای اصلی $(2, 5)$ ، که به وسیله دو خط مشخص شده است، نمی‌باشد. همچنین ماکزیم فاصله بین داده‌های شبیه سازی شده و مشاهده شده، برای سه سطح تحمل $(1/5, 1, 4/5)$ ، است. توجه می‌گردد که هر چه قدر ϵ محدودتر می‌گردد، تعداد پارامترهایی که پذیرفته می‌شوند کاهش می‌یابد. در حد وقتی $\epsilon \rightarrow 0$ و $n \rightarrow \infty$ ، به نمونه پسین واقعی نزدیک می‌شویم که هزینه بیشتری را متحمل می‌شویم. وقتی ϵ زیاد می‌شود، حجم نمونه پذیرفته شده بزرگتری داریم که بازهم هزینه افزایش می‌یابد. هر چند در عمل حجم نمونه بزرگتر مهم است، اما با مقداری خطا مواجه می‌شویم.

۲.۱.۴ مدل $MA(q)$

فرایند $MA(q)$ را بررسی می‌کنیم، که در آن $(y_k)_{k \in \mathbb{N}^*}$ یک فرایند تصادفی است و به صورت عبارت (۱.۴) تعریف می‌شود:

$$y_k = u_k + \sum_{i=1}^q \theta_i u_{k-i}, \quad (1.4)$$

که $(u_k)_{k \in \mathbb{Z}} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ ، دنباله مستقل و هم توزیع از نرمال استاندارد است. با این حال یک تحلیل بیزی را بر اساس مجموعه غیر قابل تشخیص^۱ می‌توان برای این مجموعه به‌کار برد و برآورد درستی از کمیت قابل تشخیص را بدست آورد (مارین و رابرت، ۲۰۰۷)، اما ما یک شرط قابل تشخیص استاندارد را روی این مدل تحمیل می‌کنیم، که این شرط، ریشه‌های چند جمله‌ای عبارت (۲.۴) می‌باشد.

$$Q(x) = 1 - \sum_{i=1}^q \theta_i x^i \quad (2.4)$$

آنها رویه‌های دایره واحد در یک صفحه مرکب^۲ می‌باشند. در ادامه، توزیع پیشین را توزیع یکنواخت متناظر با دامنه θ_i در نظر می‌گیریم، بخصوص زمانی که q کوچک باشد و مجموعه پارامترها را بتوان به آسانی نوشت. برای سادگی کار فرایند مورد نظر را برای زمانی که $q = 2$ است در نظر می‌گیریم. با توجه به شرط وارون پذیری باید ریشه‌های عبارت (۲.۴) داخل دایره واحد باشد. در نتیجه مثلث زیر به دست می‌آید.

$$-2 < \theta_1 < 2 \quad \theta_1 - \theta_2 > -1 \quad \theta_1 + \theta_2 < 1.$$

هرچند توزیع پیشین روی θ خیلی ساده در نظر گرفته شده و با وجود متغیرهای تصادفی نرمال، درست‌نمایی وابسته به سری $(y_k)_{1 \leq k \leq q}$ خیلی پیچیده می‌باشد چون نیاز به انتگرال‌گیری $u_0, u_{-1}, \dots, u_{-q+1}$ است. در واقع به ازای $u_k = \theta_1 u_{k-1} + \dots + \theta_q u_{k-n}$ تابع چگالی توام برای $u = (u_1, \dots, u_n)$ به صورت عبارت (۳.۴) است.

$$p(u|\theta, \mu = 0, \sigma_u^2 = 1) = (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{k=1}^n u_k^2 \right\} \quad (3.4)$$

برای اجرای تکرار یک بار ABC در این حالت نیاز هست تا :

(الف) ضرایب θ در $MA(q)$ بطور یکنواخت روی محدوده پذیرش شبیه سازی گردد.

(ب) یک دنباله مستقل و هم توزیع از $(u_k)_{-q < k \leq n}$ تولید شود.

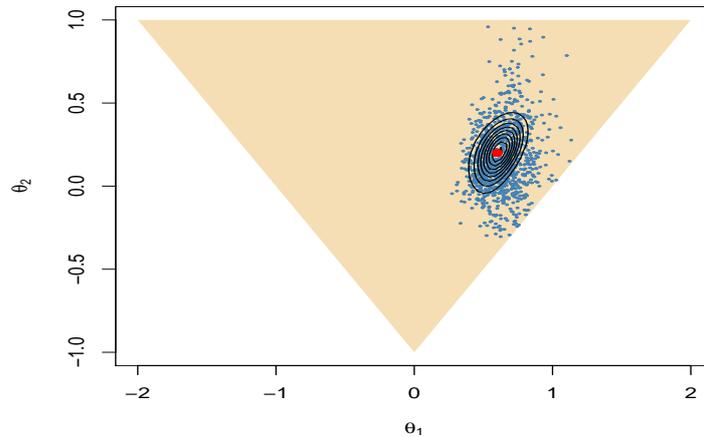
(ج) یک سری شبیه سازی شده از $(x_k)_{1 \leq k \leq n}$ ایجاد گردد.

با توجه به تمرکز و دقتی که در تجزیه تحلیل وجود دارد، فاصله می‌تواند فاصله خام بین سری‌ها باشد، یعنی $\rho^2((x_k)_{1 \leq k \leq n}, (y_k)_{1 \leq k \leq n}) = \sum_{k=1}^n (y_k - x_k)^2$ یا فاصله درجه دو بین آماره‌های خلاصه شده شبیه اتوکواریانس مرتبه اول q ، $\tau_j = \sum_{k=j+1}^n y_k y_{k-j}$ باشد که انتخابمان را برای مثال در شکل ۲.۴ نشان دادیم.

با توجه به شکل ۲.۴، نمونه ABC را با استفاده از نمونه شبیه‌سازی شده به طول ۱۰۰ و به ازای پارامتر واقعی $(\theta_1, \theta_2) = (0.6, 0.2)$ (نقطه قرمز) و همچنین سطح تحمل ϵ برابر با ۰/۱ درصد چارک نمونه‌ای فاصله مورد نظر، به پسین واقعی برازش دادیم (مجموعه سطوح با استفاده از درست‌نمایی دقیقی

^۲Complex plane

^۱Nonidentifiable

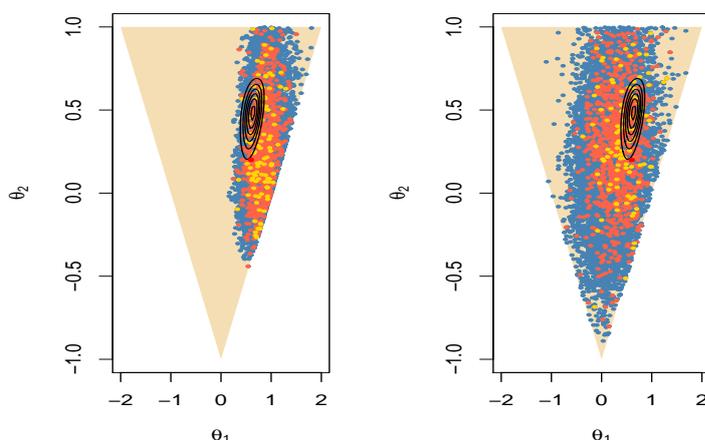


شکل ۲.۴: خروجی الگوریتم پذیرش رد بر اساس تحمل ϵ برابر با ۱٪ درصد از $N = 10^6$ داده‌های شبیه‌سازی پذیرفته‌شده

برای مدل $MA(2)$ و همچنین شبکه‌ای از مقادیر θ روی ناحیه پذیرش محاسبه شده است). مثلث، محدوده مقادیر پذیرفته شده برای θ می‌باشد. این نمودار محدوده پذیرش برای توزیع پسین واقعی را بر اساس آماره خلاصه‌شده با استفاده از اتوکواریانس در نمودار پراکنش (نقاط آبی رنگ) نیز نشان می‌دهد. ملاحظه می‌گردد که توزیع نمونه‌ای به توزیع پسین واقعی میل می‌کند، اما پسین را به طور کامل و درست ایجاد نمی‌کند. کاهش ϵ منجر به بهتر شدن چگالی پسین روی سطوح و متراکم شدن نقاط می‌گردد، اما خود این باعث افزایش هزینه به سبب اندازه نمونه بزرگتر می‌گردد.

حال به مطالعه تاثیر انتخاب فاصله ρ و تحمل ϵ روی تقریب ABC می‌پردازیم. یک نمونه 50° تایی از مدل $MA(2)$ مبتنی بر پارامترهای مشابه فوق، داریم. در ابتدا تاثیر استفاده از فاصله خام بین کل داده‌ها به جای فاصله بین اتوکواریانس بررسی می‌کنیم. شکل ۳.۴ نشان می‌دهد که فاصله خام بین داده‌های سری زمانی مشاهده شده و شبیه‌سازی شده ناکارا و تا حدودی غیر قابل تشخیص است. برای فاصله خام، پراکنش پارامترهای پذیرفته شده بعد از گام ABC خیلی پهن‌تر از فاصله دوم (با استفاده از اتوکواریانس) است، خصوصاً موقعی که با سطوحی از چگالی پسین مقایسه شود. در ادامه فقط از فاصله بین اتوکواریانس‌ها استفاده می‌کنیم.

در ادامه به انتخاب ϵ می‌پردازیم. شکل ۴.۴ نشان می‌دهد که با کاهش ϵ ، و کاهش چندک‌های تجربی فاصله شبیه‌سازی شده $\rho(s(x), s(y))$ ، تقریب را بهبود می‌بخشد، هرچند هرگز به توزیع پسین واقعی نمی‌رسیم. این خصوصاً برای پارامتر θ_2 واضح‌تر است. چگالی حاشیه‌ای نمونه‌ای ABC به وسیله برآوردگر چگالی پیش فرض در R به دست آمده است، و چگالی حاشیه‌ای واقعی با انتگرال‌گیری عددی حاصل می‌گردد. حال یک مثال پویا را بررسی می‌کنیم.



شکل ۳.۴: پراکندگی دو نمونه ABC را بر اساس فاصله خام و فاصله اتوکواریانس محاسبه شده است که با استفاده از چندک‌های متفاوت ϵ (آبی برای ۱٪، قرمز برای ۱٪، و زرد برای ۱٪) نشان داده شده است. سطوح چگالی پسین به رنگ مشکی است.

۳.۱.۴ مدل‌های زیست‌شناسی پویا

سیستم زیست‌شناسی چیست؟

بو و کاوالای (۲۰۱۱)، زیست‌شناسی^۳ را وسیله‌ای برای درک تمام زمینه‌های زندگی، از منشا و تکامل گونه‌های زیستی گرفته تا وظایفی که دارند، توصیف کرد. او همچنین سلول‌ها را به عنوان ماده‌ای که زیست می‌کنند توصیف کرد. در سیستم‌های زیستی از مدل‌های ریاضی استفاده می‌کنند تا بتوان این سیستم‌ها را توصیف کرد و همچنین رابطه بین مولفه‌های مختلف را نشان داد. این سیستم یا شبکه را باید به گونه‌ای نشان داد که همراه مجموعه‌ای از واکنش‌ها با نرخ پارامتری باشد که بتواند سرعت واکنش را نشان دهد. هدف اصلی آن است که با توجه به داده‌های مشاهده شده و مدل مورد نظر، درک خوبی از نرخ پارامتر و به لحاظ آماری فرم پسین مناسب را به دست آوریم. یکی از این مدل‌ها که رابطه بین دو گونه در یک زیست‌بوم نشان می‌دهد، مدل شکارچی-شکار^۴ یا مدل لوتکا-ولترا^۵ است.

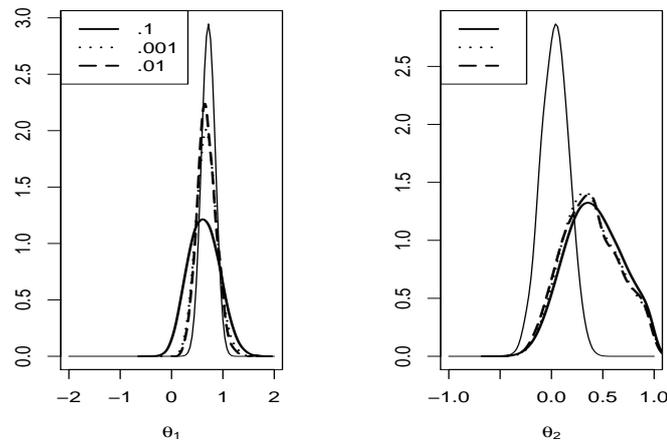
معرفی مدل

اساسی‌ترین مدل برای اثر متقابل بین گونه‌های شکارچی-شکار توسط معادلات لوتکا-ولترا ارائه شده است. برای روشن شدن مطلب به یک مثال خاص در این زمینه می‌پردازیم. فرض کنیم می‌خواهیم اثر سیاهگوش و خرگوش را بررسی کنیم. در این معادلات فرض شده که تعداد خرگوش‌ها در اثر نبود شکار به صورت نمایی رشد می‌کند و همچنین تعداد سیاهگوش‌ها در اثر کم شدن خرگوش‌ها به صورت نمایی کاهش می‌یابد. بعلاوه زمانی که سیاهگوش در سیستم وجود دارد، جمعیت خرگوش‌ها به صورت نمایی

^۵Lotka-Volterra

^۳system biology

^۴predator-prey



شکل ۴.۴: توزیع نمونه‌ای ABC را با استفاده از چندک‌های متفاوت نقطه نقطه برای ۱۰٪، خط تیره برای ۱٪ و خط توپر برای ۱٪ درصد، که در مقایسه با چگالی حاشیه‌ای واقعی نشان داده شده است. داده‌ها مشابه نمودار ۳.۴ است.

کاهش می‌آید. به همین نسبت جمعیت سیاه‌گوش‌ها با توجه به نرخ وابستگی که با جمعیت خرگوش‌ها دارد کاهش یا افزایش خواهد یافت. این پویایی در چرخه رشد و زوال ادامه دارد. اگر $u_1(t)$ برابر تعداد سیاه‌گوش‌ها در زمان t ، و $u_2(t)$ را برابر تعداد خرگوش‌ها در زمان t باشد، آنگاه یک چارچوب ساده برای شرح پویایی اثر متقابل دو گونه به صورت معادلات دیفرانسیلی زیر بیان می‌گردد.

$$\begin{cases} \frac{du_1}{dt} = -\alpha_1 u_1 + \beta_1 u_1 u_2 \\ \frac{du_2}{dt} = \alpha_2 u_2 - \beta_2 u_1 u_2 \end{cases} \quad (4.4)$$

با توجه به عبارت (۴.۴)، $\frac{du_1}{dt}$ و $\frac{du_2}{dt}$ بیانگر نرخ رشد دو جامعه نسبت به زمان t است، و α_j و β_j برای $j = 1, 2$ پارامترهای مثبت مدل هستند. یعنی:

- $\alpha_1 u_1$: میزان از دست دادن شکارچیان بعلت مرگ طبیعی یا کوچ کردن است.
- $\beta_1 u_1 u_2$: رشد جمعیت شکارچی است.
- $\beta_2 u_1 u_2$: نرخ که شکارچیان و طعمه با هم روبرو می‌شوند
- $\alpha_2 u_2$: رشد (تولید مثل) نمایی طعمه

از منظر زیست‌شناسی، این مدل دارای نقیصه آشکار است، و آن هم دیگر شکارچیان است که بر روی جمعیت طعمه، در نظر گرفته نمی‌شود. یعنی مدلی را می‌طلبند که بیان‌کننده اثر متقابل بیشتری در این زیست‌گاه^۶ باشد، به طوری که در معادله دوم عبارت (۴.۴)، عبارات بیشتری از $-\beta_j u_j u_2$ برای

^۶habitats

$J, 3,000, J$ داشته باشیم. $J-1$ ، تعداد شکارچیان است که طعمه را شکار می‌کنند. در واقع می‌توان سیستمی را فرض کرد که برای هر شکارچی یک معادله وجود دارد و برای هر طعمه یک معادله وجود دارد به طوری که یکدیگر را در یک زیست‌گاه مصرف می‌کنند. اما چیزی که باعث می‌شود معادله سیاهگوش منحصر بفرد و فقط تحت تاثیر جمعیت خرگوش‌ها باشد این است که فرض کنیم دو نوع خرگوش داریم: آن‌هایی که توسط سیاهگوش مصرف می‌شوند و آن‌هایی که نمی‌شوند. یعنی معادلات را برای جمعیت کل خرگوش‌ها در دو معادله افراز کنیم. یکی از دو معادله پویایی جمعیت خرگوش‌هایی است که توسط سیاهگوش مصرف می‌شوند را کنترل می‌کند و معادله دیگر برای دیگر شکارچیان که خرگوش را شکار می‌کنند. معادله اول به جمعیت شکارچیان دیگر وابسته نیست و دقیقاً فرمی به شکل عبارت (۴.۴) دارد. این دو معادله بهم مرتبط هستند. فرض می‌کنیم که اثر رقابت بین خرگوش‌ها در مقایسه با اثر تولد و مرگ بر روی جمعیت، ناچیز است. چنین فرضیه‌ای اساساً گرایش به معادلات لوتکایی می‌باشد. بنابراین برای نشان دادن تاثیر شکارچیان دیگر، فقط باید در معادلات لوتکا-ولترا تغییراتی ایجاد کرد. البته در حال حاضر نمی‌توان تعداد طعمه‌ها را که به وسیله شکارچیان شکار می‌گردند را تعیین نمود. در واقع اندازه‌گیری تعداد خرگوش‌ها در یک زیست‌گاه کار دشواری است.

یکی دیگر از ضعف‌های زیست‌شناسی این مدل این است که تعداد خرگوش‌ها بعلت عدم وجود شکارچی، بدون حد افزایش می‌یابد. این رشد تا زمانی که منابع غذایی موجود باشد ادامه دارد. برای حل این مشکل، عبارات دیگری به سمت راست معادله عبارت (۴.۴) اضافه می‌گردد، که شامل قدرت جمعیت گونه در سمت چپ معادله عبارت (۴.۴) است به طوری که این رفتار کنترل گردد. به جای استفاده از این روش، از دستگاه معادلات استفاده می‌کنیم که به عنوان یک مدل مید تنها زمانی که شرایط طوری است که گونه از بین نمی‌رود، استفاده می‌کنیم. این شرایطی است که برای چند صد سال بر روی سیستم سیاهگوش-خرگوش قابل اجرا بوده است، بنابراین این مدل برای سیستم سیاهگوش-خرگوش مناسب است. به طور مشابه برای چندین جفت گونه دیگر به کار می‌رود. از جنبه آمار، این مدل برای برخی گونه‌ها به عنوان مدل مناسب نیست. آن هم زمانی است که معادلات دارای ساختاری پایدار نباشند: تغییر کوچک در مقدار پارامتر منجر به تغییر اساسی در رفتار معادلات می‌گردد. در نتیجه باعث می‌شود که یا از معادلات صرف نظر کنیم یا آن را برای به دست آوردن یک سیستم بهتر تغییر دهیم. هنگامی که این ناپایداری سبب دشواری در برازش مدل گردد، می‌توان از معادلاتی استفاده کرد که پارامتر را برآورد کند و پیش‌گوهایی را بدست آورد که در ادامه به آن خواهیم پرداخت. نمی‌توان پنداشت که ناپایداری در مدل باعث غیر واقعی بودن مدل گردد، چون جهان پر از پدیده‌ها همراه با پارامترهای بسیار حساس است.

مدل تصادفی

بویز و همکاران (۲۰۰۷) یک نوع مدل تصادفی منظم را در حالت کلی به صورت عبارت (۵.۴) نشان دادند، که X نشان‌دهنده گونه یا ژن مورد نظر است. همچنین $n, 1,000, n$ برای n مدل تصادفی است و p نشان‌دهنده ماتریس ضرایب پیشین و q نشان‌دهنده ماتریس ضرایب پسین است.

$$R_i = p_{i1}X_1 + \dots + p_{iu}X_u \longrightarrow q_{i1}X_1 + \dots + q_{iu}X_u \quad (5.4)$$

دارن ویلکینسون (۲۰۰۷) مدل لوتکا-ولترا را به لحاظ مدل تصادفی عبارت (۵.۴) اینگونه توضیح داد، که مدل را با توجه به (۴.۴) مجموعه‌ای از واکنش‌های جفتی در نظر گرفت، یعنی

$$R_1 : u_2 \xrightarrow{\alpha_2} 2u_2 \quad (۶.۴)$$

$$R_2 : u_2 + u_1 \xrightarrow{\beta} 2u_1 \quad (۷.۴)$$

$$R_3 : u_1 \xrightarrow{\alpha_1} \emptyset. \quad (۸.۴)$$

با توجه به مدل لوتکا-ولترا، عبارت (۶.۴) بیانگر تولید مثل شکار با نرخ α_2 است و عبارت (۷.۴) بیانگر تولید شکارچی و اثر متقابل بین شکار و شکارچی با نرخ β است. شکارچی، شکار را مصرف می‌کند و این به طور مستقیم تاثیر در نرخ تولید شکارچی دارد. عبارت (۸.۴)، بیانگر مرگ شکارچی به علت طبیعی با نرخ α_1 است. مرتبط با پارامترهای مدل، تابع مخاطره یا تابع نرخ تصادفی مدل تعریف می‌شود. با توجه به جدول ۱.۴ تابع مخاطره مرتبط با نرخ واکنش مورد نظر نشان داده شده است.

با توجه به آنچه که بیان شد، فرایند مارکف را با هسته (چگالی) انتقال رام نشدنی که بر اساس هسته

جدول ۱.۴: با توجه به عبارات (۶.۴)، (۷.۴) و (۸.۴)، منظور از پیشین، ضرایب سمت راست عبارات مذکور با توجه به u_1 و u_2 است. همین حالت برای پسین برقرار است. اختلاف پسین از پیشین، مخاطره مورد نظر را می‌دهد.

مخاطره	پسین	پیشین
$\alpha_2 u_2$	۲ ۰	۱ ۰
$\beta u_1 u_2$	۰ ۲	۱ ۱
$\alpha_1 u_1$	۰ ۰	۰ ۱

مدل لوتکا-ولترا می‌باشد، را در نظر می‌گیریم. استنباط برای چنین مدلی چالشی بزرگ است. با این حال شبیه سازی از مدل امکان پذیر است. برای مثال می‌توان از الگوریتم گلیپس (گلیپس ۱۹۹۷) استفاده کرد. ذکر این مثال، یک کاربرد طبیعی برای ABC می‌باشد. با توجه به پارامترهای مدل که $\theta = \{\alpha_1, \beta, \alpha_2\}$ هست و با توجه به هسته رام نشدنی و فرم نامشخص درست‌نمایی مدل لوتکا-ولترا که در پی آن فرم غیر بسته پسین، نیاز هست تا از روش ABC برای شبیه سازی پارامترهای مدل لوتکا-ولترا استفاده گردد. به عنوان مثال، بویز و همکاران (۲۰۰۸)، فرم غیر بسته تابع درست‌نمایی را برای مدل لوتکا-ولترا نشان دادند و در نتیجه سخت بودن استنباط درست‌نمایی حاصله از آن را بررسی کردند. بویز و همکاران (۲۰۰۸) و پرانگل (۲۰۱۱) نشان دادند بر اساس مدل تصادفی لوتکا-ولترا

تابع احتمال آن به صورت عبارت (۹.۴) در می‌آید.

$$p(U_{t+\Delta t} = (z_1, z_2) | U_t = (u_1, u_2)) = \begin{cases} 1 - (\alpha_1 u_1 + \beta u_1, u_2 + \alpha_2 u_2) + o(\Delta t) & \text{اگر } z_2 = u_2 \text{ و } z_1 = u_1 \\ \alpha_1 u_1 + o(\Delta t) & \text{اگر } z_2 = u_2 \text{ و } z_1 = u_1 + 1 \\ \beta u_1, u_2 + o(\Delta t) & \text{اگر } z_2 = u_2 + 1 \text{ و } z_1 = u_1 - 1 \\ \alpha_2 u_2 + o(\Delta t) & \text{اگر } z_2 = u_2 - 1 \text{ و } z_1 = u_1 \\ o(\Delta t) & \text{در غیر این صورت} \end{cases} \quad (9.4)$$

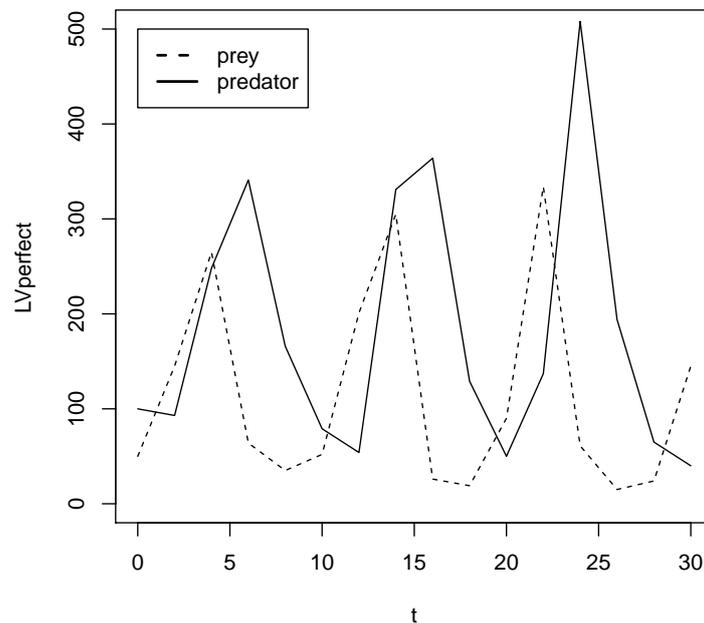
با توجه فرم تابع احتمال عبارت (۹.۴)، که به صورت چند بخشی در آمده است، مقدار گونه مورد نظر در دو زمان متفاوت نشان داده است. در نتیجه نمی‌توان فرم بسته‌ای برای درستنمایی آن به دست آورد. با توجه به مدل تصادفی ویلکینسون (۲۰۰۶) برای مدل لوتکا-ولترا، او با در نظر گرفتن عبارت (۹.۴) به عنوان تابع چگالی احتمال و روش گیلیپس، فرم تابع درستنمایی را برای یک مدل تصادفی در حالت کلی به صورت زیر به دست آورد.

$$\pi(u(i) | \theta) = \left\{ \prod_{i=0}^{T-1} h_{ki}[u(t_{i-1}), \theta_{ki}] \right\} \times \exp \left\{ - \int_0^T h_0[u(t), \theta] dt \right\} \quad (10.4)$$

که $h_k(u, \theta_k) = \theta_k g_k(y)$ به ازای $k = 1, \dots, \nu$ تابع مخاطره^۷ است و $t_i \equiv i$ است. در این صورت فرم درستنمایی برای مدل لوتکا-ولترا در حالت مدل تصادفی برای $k = 1, 2, 3$ برقرار می‌شود که فرم غیر بسته همراه با انتگرال گیری رام نشدنی می‌باشد. مدل مورد بررسی از دو مولکول تشکیل شده است. متغیر زمان بر اساس متغیر u را با (u_{1t}, u_{2t}) مشخص می‌کنیم. سه نوع عکس العمل وجود دارد: نوع ۱ برای پیدایش مولکول، نوع ۲ برای از بین رفتن مولکول و نوع ۳ برای عکس العملی (واکنش) که بین دو مولکول حذف شده و مولکول اضافه شده وجود دارد. برای ارتباط این مثال با مدل لوتکا-ولترا، مولکول نوع ۱ شکار و مولکول نوع ۲ را شکارچی در نظر می‌گیریم. از کاربرد دیگر این مدل با توجه به عبارت (۵.۴)، می‌توان به استفاده آن در علم ژنتیک نام برد. برای این مثال خاص از بسته *smfsb* در R استفاده می‌کنیم. در این بسته دو تابع *simTs* و *stepLV* قرار دارند، تا براساس N پارامتر شبیه سازی شده، N تحقق x را بدست آوریم. در این بسته از *LVperfect* به عنوان داده واقعی استفاده می‌کنیم و مبنای مقایسه را بر اساس آن قرار می‌دهیم. در بسته *smfsb* مقدار اولیه برای *LVperfect* در فاصله زمانی $t = [0, 30]$ در نظر گرفته شده است، که در نمودار ۵.۴ آمده است. سپس با استفاده از الگوریتم گیلیپس از روی این داده اولیه، به تعداد دلخواه داده شبیه سازی شده است. ما این داده‌های شبیه سازی شده را به عنوان داده واقعی مان در نظر می‌گیریم. سپس استنباط روی آن را انجام می‌دهیم.

همچنین مقدار واقعی پارامتر را $\theta = \{1, 0.005, 0.16\}$ در نظر می‌گیریم. در این صورت با توجه به N تحقق شبیه سازی شده و مقایسه با *LVperfect*، N مقایسه $\|y - x\|$ با استفاده از نرم‌های مختلف،

^۷function hazard

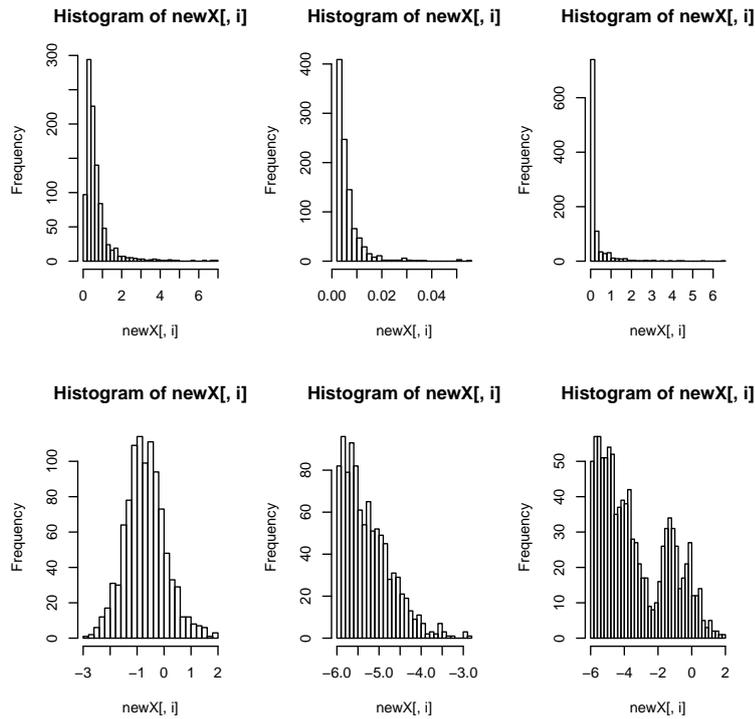


شکل ۵.۴: نمودار داده $LV_{perfect}$ نسبت به زمان t است که خط چین بیانگر داده شکارچی و خط پر بیانگر داده شکار است.

خواهیم داشت. در نهایت θ بی را می‌پذیریم که نرم مورد نظر از ϵ کمتر باشد. انتخاب و تعیین تحمل ϵ در فصل ۳ گفته شده است. در اینجا ϵ را از یک چندک توزیع فاصله‌ای انتخاب می‌کنیم. در نهایت ۱۰۰۰ θ پذیرفته می‌گردد. به علت حجم محاسبات بالا، هیستوگرام مقادیر پارامترهای پذیرفته شده را یک بار بدون بسته *parallel* و یک بار با استفاده از بسته *parallel* رسم می‌کنیم.

با توجه به شکل‌های ۶.۴ و ۷.۴، برای اینکه شکل توزیع پسین تقریبی واضح‌تر باشد از $\log(\theta)$ استفاده می‌کنیم. واضح است که اگر از یک نرم مناسب استفاده کنیم و $\epsilon \rightarrow 0$ میل کند آنگاه توزیع مقادیر پذیرفته شده به توزیع پسین مطلوب میل می‌کند. هرچند انتخاب ϵ خیلی کوچک، منجر به نرخ رد بالاتر می‌گردد. وقتی بعد y زیاد می‌شود آنگاه غیر واقعی است که انتظار داشته باشیم تا یک هماهنگی نزدیکی بین مولفه‌های x و y وجود داشته باشد. در نتیجه در عین پرهزینه بودن، برای بعد بالا این امر تقریباً غیرممکن می‌گردد. پس به ناچار از آماره خلاصه شده و در صورت وجود از آماره بسنده استفاده می‌کنیم. در مثالی که گذشت بعد مشاهدات ۱۶ است و با توجه به ۲ متغییره بودن مدل، در واقع بعد ۳۲ است. برای کاهش بعد در ادامه از میانگین و واریانس و اتو کواریانس مشاهدات استفاده می‌کنیم. نتیجه آن را در شکل ۸.۴ می‌بینیم.

شکل ۸.۴ توزیع پسین تقریبی را نشان می‌دهد، که خط قرمز بیانگر مقدار واقعی پارامتر است. همان‌طور که مشخص است مقادیر پذیرفته شده حول مقدار واقعی است (شکل ۸.۴ را ببینید). نمودار به دست آمده از شکل ۸.۴ خیلی شبیه نمودار با استفاده از کل داده‌ها می‌باشد. در اینجا کاهش بعد زیاد نیست. در واقع از ۳۲ مشاهده واقعی به ۹ مشاهده کاهش یافته است. در نتیجه عملکرد آن با استفاده از آماره

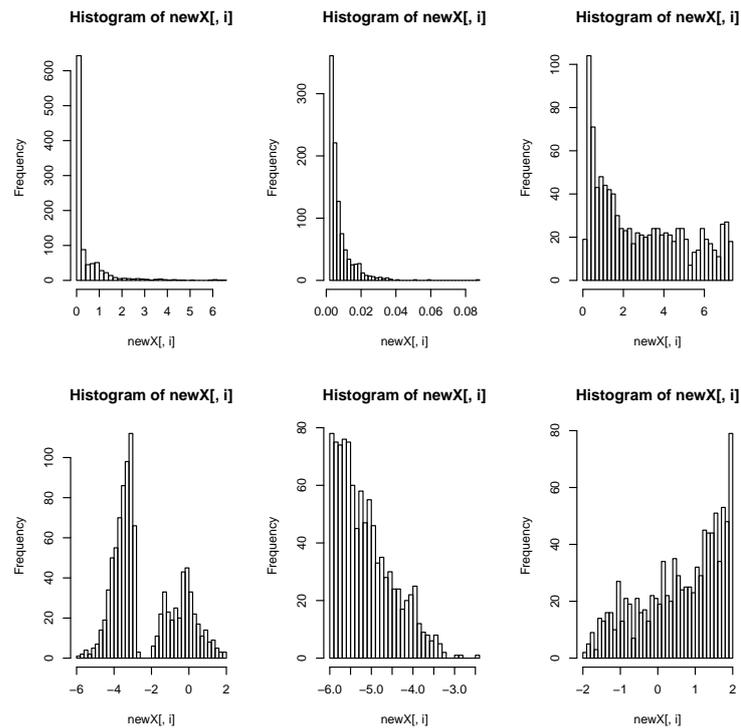


شکل ۶.۴: نمودار هیستوگرام پارامترهای شبیه سازی شده بر اساس θ و $\log(\theta)$

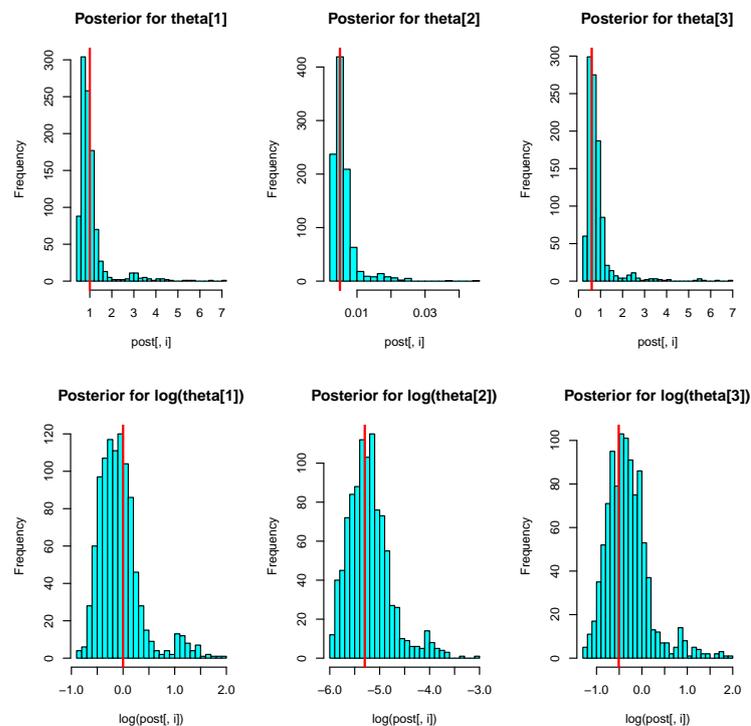
خلاصه شده را در زمان محاسبه شده با استفاده از ABC می توان دید. جدول ۲.۴ زمان محاسبه شده را بر اساس استفاده از بسته *parallel* و آماره خلاصه شده نشان می دهد. بنابراین کاهش بعد در ابعاد بالاتر خیلی ضروری به نظر می رسد، چراکه با افزایش بعد هزینه نیز افزایش می یابد. در نتیجه مهم ترین جنبه در کاهش زمان استفاده از آماره خلاصه شده بهینه (میانگین پسین) یا تقریبی از آماره خلاصه شده است.

جدول ۲.۴: زمان محاسبه شده برای تقریب پسین

تقریب پسین	زمان
بدون بسته <i>parallel</i>	۵۰۳.۱۱
با بسته <i>parallel</i>	۴۶۵.۶۵
با <i>parallel</i> و آماره خلاصه شده میانگین و واریانس و اتوکوریانس	۴۱۱.۶۸



شکل ۷.۴: نمودار هیستوگرام پارامترهای شبیه سازی شده بر اساس θ و $\log(\theta)$ با استفاده از بسته *parallel*



شکل ۸.۴: نمودار هیستوگرام پارامترهای شبیه سازی شده بر اساس θ و $\log(\theta)$

نتیجه‌گیری و آینده تحقیق

محاسبات بیز تقریب برای رده گسترده‌ای از مدل‌هایی که در دسترس نیستند، استنباط را ممکن می‌سازد. این روش در هر دو زمینه نظری و کاربردی مجموعه‌ای از توجهات را به خود جلب کرده است. در تعمیم‌های اخیر، با توجه به کالبدن ABC، در بسیاری از حالات منجر به تقریبی و کارایی بسیار خوبی شده است. کارایی روش در خروجی به دست آمده از الگوریتم‌های مختلف ABC می‌تواند به خوبی تاثیرگذار باشد. اخیراً روش ABC هم توسط آماردانان و هم پزشکان تحت بررسی دقیق قرار گرفته است. از این رو، هدف بهبود عملکرد روش ABC است. در حقیقت این بهتر شدن، پاسخی به ملاحظات برخی از آماردانان، درباره اعتبار رویکرد، و پزشکان درباره کارایی بالاتر روش است. در این میان بعضی از مسایل حل نشده باقی می‌ماند از جمله:

- نتایج همگرایی به دست آمده تاکنون غیر عملی هستند، چراکه نیاز هست تا سطح تحمل به دست آمده کوچک گردد (به سمت صفر میل کند) یا اندازه نمونه به بینهایت میل کند، که این خود هزینه بسیاری را در پی دارد. به دست آوردن خطای دقیق مرزها برای سطح تحمل مثبت و انداز نمونه متناهی هم در اجرا و هم در ارزیابی عملکرد روش و هزینه نتایج بهتری را به دنبال دارد.
- با اینکه ABC غالباً به عنوان یک روش همگرایی و تقریبی در استنباط بیزی بیان می‌شود اما آن می‌تواند به خودی خود به عنوان یک تکنیک استنباطی بیان شود، در نتیجه تحلیل آن سر راست است. در حال حاضر به عنوان یک روش استنتاجی غیر مستقیم به کار می‌رود.
- اگرچه ABC تاحدی مساله تقریب را حل کرده است اما با افزایش بعد مجموعه خطاهایی که در فصل ۳ اشاره شد، نیز افزایش می‌یابد و این کارایی را کاهش و با افزایش هزینه و زمان مواجه می‌گردد. با این حال روش ABC برای نمونه‌های کوچک‌تر بسیار سریع و راحت و ارزان است.
- رابطه استنباط مبتنی بر ABC با روش‌های تقریبی دیگر مانند استنباط بیزی، variational تا کنون انجام نشده است مقایسه و درهم آمیختن این روش‌ها منجر به یک تحقیق پژوهشی خوب می‌گردد (جاکولا و همکاران، ۲۰۰۰).
- انتخاب و ساختن آماره خلاصه شده تاکنون به صورت تجربی صورت گرفته است. بر اساس یک رویکرد خودکار و بر اساس اصول تحلیل داده‌ها^۸ و تقریب بسندگی، این امر می‌تواند بسیار جذاب‌تر و قانع‌کننده‌تر به ویژه در مجموعه‌های پیچیده و غیراستاندارد باشد.

^۸Data analysis

- روش لزوماً با محدودیت‌هایی از جمله مجموعه داده‌های بزرگ و مدل‌های پیچیده مواجه هست، به طوری که شبیه‌سازی به خودی خود به یک امر غیر ممکن تبدیل گردد. در نتیجه به سرعت تکنیک کاهش بعد به طور مستقیم برای شبیه‌سازی آماره‌های خلاصه شده نیاز خواهد بود.
- یک جنبه خاص از تحلیل بیزی انتخاب مدل از چندین مدل به طور هم‌زمان است که پیچیدگی محاسباتی را نیز داراست (رابرت، ۲۰۰۱؛ مارین و رابرت، ۲۰۱۰). بر این اساس می‌توان از ABC برای حل این مساله کمک گرفت. در این صورت پارامترهای هر مدل، استنباط را برای مدل مربوطه با شاخص M ، که مرتبط با توزیع پیشین $\pi(M = m)$ است، به طور هم‌زمان از $(m = 1, \dots, M)$ توزیع پیشین به شرط پارامتر مربوطه روی فضای پارامتر Θ_m انجام می‌دهد. انتخاب بین این مدل‌ها با محوریت توزیع پسین از M ، یک چالش محاسباتی است که ABC به حل آن می‌پردازد (مارین و همکاران، ۲۰۱۲؛ گرلود و همکاران، ۲۰۰۹؛ لیونبرگر و همکاران، ۲۰۱۰؛ رابرت و همکاران، ۲۰۱۰؛ رابرت و همکاران، ۲۰۱۱).

پیوست آ

نمادها و تعریفها

تعریف ۱. فضای برداری^۱

مجموعه‌ای است مانند V روی میدان F ، که مجهز به دو عملگر، جمع برداری و ضرب عددی است. همچنین دارای γ خاصیت زیر است:

- به‌ازای هر x و y در V ، $x + y = y + x$
- به‌ازای هر x و y و z در V ، $(x + y) + z = x + (y + z)$
- عنصری مانند 0 در V هست، به‌قسمی که به‌ازای هر x در V ، $x + 0 = x$ و $0 + x = x$
- به‌ازای هر x در V ، عنصری مانند $-x$ در V هست به‌قسمی که $x + (-x) = 0$
- به‌ازای هر x ، $x \in V$ ، $1x = x$
- به‌ازای هر $a, b \in \mathbb{R}$ و $x \in V$ ، $a(bx) = (ab)x$
- به‌ازای هر $a, b \in \mathbb{R}$ و $x, y \in V$ ، $a(x + y) = ax + ay$ و $(a + b)x = ax + bx$

تعریف ۲. فضای L_p

فضای L_p به‌عنوان فضایی از توابع قدرطلق توان p ام انتگرال لبگ تعریف می‌شود. به‌طور کلی، فرض کنید که $1 \leq p < \infty$ ، و فضای اندازه (S, Σ, μ) باشد. مجموعه‌ای از تمام توابع اندازه‌پذیر از S به C یا R را در نظر می‌گیریم. در این صورت توان p ام قدر مطلق دارای انتگرال متناهی است به طوری که

$$\|f\|_p = \left(\int_S |f|^p d\mu \right)^{\frac{1}{p}} < \infty.$$

^۱Vector space

مجموعه‌ای از چنین توابعی، فضای برداری را تشکیل می‌دهند. با دو عملگر طبیعی و برای هر λ داریم

$$(f + g)(x) = f(x) + g(x)$$

$$(\lambda f)(x) = \lambda f(x).$$

همچنین در فضای L_p نامساوی زیر را نیز داریم.

$$\|f + g\|_p \leq 2^{p-1}(|f|^p + |g|^p)$$

تعریف ۳. زنجیر مارکوف

یک زنجیر مارکوف $\{\theta^{(t)}\}$ دنباله‌ای از متغیرهای تصادفی وابسته $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(t)}, \dots$ است، به طوری که توزیع احتمال $\theta^{(t)}$ به شرط مفروض بودن متغیرهای قبلی فقط به $\theta^{(t-1)}$ وابسته است.

تعریف ۴. زنجیر تحویل‌ناپذیر

تحویل‌ناپذیری به این معنی است که هسته $k(\cdot|\cdot)$ اجازه حرکت بر روی تمام وضعیت‌های فضای حالت زنجیر را دارد. یعنی، صرف‌نظر از مقدار اولیه $\theta^{(0)}$ ، دنباله $\{\theta^{(t)}\}$ با احتمال مثبت به هر ناحیه از فضای حالت می‌تواند سرریزند.

تعریف ۵. زنجیر بازگشتی

بازگشتی بودن به این معنی است که زنجیر به هر مجموعه دلخواه غیر قابل اغماض، به دفعات نامتناهی باز خواهد گشت.

تعریف ۶. زنجیر آرگودیک

در زنجیرهای بازگشتی، توزیع مانا یک توزیع حدی نیز هست. به این معنی که توزیع حدی $\theta^{(t)}$ به ازای هر مقدار اولیه $\theta^{(0)}$ ، $\pi(\theta|\cdot)$ است. این ویژگی را آرگودیک بودن می‌نامند. به عبارت دیگر ویژگی آرگودیک بودن زنجیر، با استقلال نسبت به شرایط اولیه، معادل است.

پیوست ب

کد برنامه R برای اجرای فصل دوم و چهارم

دستورات لازم برای بازتولید نتایج مثال اول در فصل دوم و مثال دوم و سوم در فصل چهارم در این پیوست گزارش شده است.

مثال اول فصل دوم

• دستورات مربوط به شکل ۱.۲ و برای $N = ۲۰$ داده واقعی و شبیه‌سازی شده به صورت زیر می‌باشد.

```
make_ecdf <- function(X) {  
  list(x=c(0,X,10), y=c(0:5/5,1))  
}  
  
dist <- function(X,Y) {  
  sqrt(sum((X-Y)^2))  
}  
  
set.seed(123)  
mu0 <- runif(1,4,6)  
Xobs <- sort(rnorm(5,mu0,1))  
ecdf_obs <- make_ecdf(Xobs)  
  
N <- 1E4  
mu_sim <- runif(N,0,10)  
Xsim <- sapply(1:N,  
  function(i){  
    sort(rnorm(5,mu_sim[i],1))
```



```

    })
Xsim <- t(Xsim)
dist_sim <- sapply(1:N,
                  function(i){
                    sqrt(sum((Xobs-Xsim[i,])^2))
                  })
acc_sim <- (dist_sim < 2)
plot_col <- -6*acc_sim+8

Sobs <- mean(Xobs)
Ssim <- rowMeans(Xsim)
dist_sim2 <- abs(Ssim - Sobs)

##PLOTS
##Plot observed vs first 20 simulated data sets
par(mfrow=c(1,2))
plot(ecdf_obs$x, ecdf_obs$y,
     type='s', xlim=c(0,10), lwd=3,
     xlab='Y', ylab='ecdf', main='Data')
for (i in which(!acc_sim[1:20])) {
  ecdf_sim <- make_ecdf(Xsim[i,])
  lines(ecdf_sim$x, ecdf_sim$y,
        type='s', col='grey')
}
for (j in 1:2) {
  i <- which(acc_sim)[j]
  ecdf_sim <- make_ecdf(Xsim[i,])
  lines(ecdf_sim$x, ecdf_sim$y,
        type='s', col=c('red','blue')[j], lwd=c(1.4,1)[j])
}
plot(mu_sim[1:20], dist_sim[1:20],
     xlim=c(0,10), ylim=c(0,14), cex=1, col=plot_col[1:20], pch=20,
     xlab="theta", ylab="Distance")
abline(h=2, col='orange', lwd=1)

```

- دستورات مربوط به شکل ۲.۲ برای $N = 1000$ داده واقعی و شبیه‌سازی شده است. در این مثال از ϵ مختلف استفاده شده است، که به صورت زیر می‌باشد.

```
par(mfrow=c(4,2), mar=c(2, 4, 2, 2) + 0.1)
for (hh in quantile(dist_sim[1:1E3],c(0.01,0.05,0.1,0.25))) {
  acc_sim <- (dist_sim < hh)
  plot_col <- -6*acc_sim+8
  dens <- density(mu_sim[1:1E3][acc_sim[1:1E3]])
  plot(dens$x, dens$y,
       type='l', xlim=c(2,8), ylim=c(0,1.5),
       xlab='', ylab='Density')
  myx <- seq(0,10,length.out=100)
  myy <- dnorm(myx, mean(Xobs), sd=sqrt(1/5))
  lines(myx, myy, lty=2)
  #points(mu_sim[acc_sim], rep(0,sum(acc_sim)), pch="|")
  plot(mu_sim[1:1E3], dist_sim[1:1E3],
       xlim=c(0,10), ylim=c(0,14), pch=16, cex=0.5, col=plot_col,
       xlab='', ylab="Distance")
  abline(h=hh, col='orange', lwd=1)
}
```

- دستورات مربوط به شکل ۳.۲، در ادامه شکل ۲.۲، با استفاده از آماره خلاصه شده $S(y) = \bar{y}$ ، به صورت زیر می‌باشد.

```
par(mfrow=c(4,2), mar=c(2, 4, 2, 2) + 0.1)
for (hh in quantile(dist_sim2[1:1E3],c(0.01,0.05,0.1,0.25))) {
  acc_sim <- (dist_sim2 < hh)
  plot_col <- -6*acc_sim+8
  dens <- density(mu_sim[1:1E3][acc_sim[1:1E3]])
  plot(dens$x, dens$y,
       type='l', xlim=c(2,8), ylim=c(0,1.5),
       xlab='', ylab='Density')
  myx <- seq(0,10,length.out=100)
  myy <- dnorm(myx, mean(Xobs), sd=sqrt(1/5))
  lines(myx, myy, lty=2)
  #points(mu_sim[acc_sim], rep(0,sum(acc_sim)), pch="|")
}
```

```

plot(mu_sim[1:1E3], dist_sim2[1:1E3],
      xlim=c(0,10), ylim=c(0,6), pch=16, cex=0.5, col=plot_col,
      xlab='', ylab="Distance")
abline(h=hh, col='orange', lwd=1)
}

```

مثال دوم فصل چهارم

- دستورات مربوط به شکل ۲.۴ و با داده اولیه $n = 10^6$ و داده شبیه‌سازی شده است. آماره خلاصه شده اتوکواریانس است. از روش RS-ABC استفاده شده است.

```

library(mnormt)
ma2like=function(theta)
{
n=length(y)
sigma=toeplitz(c(1+theta[1]^2+theta[2]^2,theta[1]
+theta[1]*theta[2],theta[2],rep(0,n-3)))
if (theta[2]<1 & theta[2]>-1 & theta[1]>-theta[2]-1 & theta[1]<theta[2]+1)
dmnorm(y,rep(0,n),sigma) else 0
}

n=100
y=as.vector(arima.sim(n=n,list(ma=c(0.6,0.2))))
acy=c(sum(y[1:(n-1)]*y[2:n]),sum(y[1:(n-2)]*y[3:n]))

acysim=matrix(0,10^4,2)
for (j in 1:10^4)
{
theta=runif(2,c(-2,-1),c(2,1))
while((sum(theta)<{-1})||(diff(theta)<{-1})) theta=runif(2,c(-2,-1),c(2,1))
ysim=as.vector(arima.sim(n=n,list(ma=c(theta[1],theta[2]))))
acysim[j,]=c(sum(ysim[1:(n-1)]*ysim[2:n]),sum(ysim[1:(n-2)]*ysim[3:n]))
}
Sigma1=cov(acysim)
ISigma1=solve(Sigma1)

```

```

Nsimu=10^6
dede=rep(0,Nsimu)
theta=matrix(0,Nsimu,2)

for (i in 1:Nsimu)
{
theta[i,]=runif(2,c(-2,-1),c(2,1))
while((sum(theta[i,])<{-1})||(diff(theta[i,])<{-1})) theta[i,]
=runif(2,c(-2,-1),c(2,1))
ysim=as.vector(arima.sim(n=n,list(ma=c(theta[i,1],theta[i,2])))
acysim=c(sum(ysim[1:(n-1)]*ysim[2:n]),sum(ysim[1:(n-2)]*ysim[3:n]))
dede[i]=t(acy-acysim)%*%ISigma1%*(acy-acysim)
print(i)
}

ma2like=function(theta)
{
n=length(y)
sigma=toeplitz(c(1+theta[1]^2+theta[2]^2
,theta[1]+theta[1]*theta[2],theta[2],rep(0,n-3)))
if (theta[2]<1 & theta[2]>-1 & theta[1]>-theta[2]-1 & theta[1]<theta[2]+1)
dmnorm(y,rep(0,n),sigma) else 0
}

len=100
A=seq(-2,2,length=len)
B=seq(-1,1,length=len)
Z=expand.grid(A,B)
Z=as.matrix(Z)
resi=rep(0,len*len)
for (i in 1:(len*len)) resi[i]=ma2like(Z[i,])
li=quantile(dede,0.001)
plot(0,0,cex=.4,pch=19,col="wheat"
,xlab=expression(theta[1]))

```

```
,ylab=expression(theta[2]),xlim=c(-2,2),ylim=c(-1,1))
polygon(c(-2,0,2,-2),c(1,-1,1,1),col="wheat",border=F)
points(theta[dede<=li,],col="steelblue",pch=19,cex=.4)
contour(A,B,matrix(resi,len,len),add=TRUE,nlev=10,drawlabels=FALSE)
points(0.6,0.2,pch=19,cex=1,col="red")
```

• در ادامه شکل ۲.۴، با استفاده از RS-ABC و MCMC-ABC، شکل ۳.۴ کشیده شده است.

```
n=50
obs=as.vector(arima.sim(n=n,list(ma=c(0.6,0.2))))
acobs=c(sum(obs[1:(n-1)]*obs[2:n]),sum(obs[1:(n-2)]*y[3:n]))

ysim=matrix(0,10000,n)
for (j in 1:10000)
{
ysim[j,]=as.vector(arima.sim(n=n,list(ma=c(theta[j,1],theta[j,2]))))
}
Sigma2=cov(ysim)
ISigma2=solve(Sigma2)

Nsimu=1000000
dede1=dede2=rep(0,Nsimu)
beaumont=matrix(0,Nsimu,2)

for (i in 1:Nsimu)
{
ysim=as.vector(arima.sim(n=n,list(ma=c(theta[i,1],theta[i,2]))))
acysim=c(sum(ysim[1:(n-1)]*ysim[2:n]),sum(ysim[1:(n-2)]*ysim[3:n]))
beaumont[i,]=acysim-acobs
dede1[i]=t(beaumont[i,])%*%ISigma1%*%beaumont[i,]
dede2[i]=t(obs-ysim)%*%ISigma2%*(obs-ysim)
print(i)
}

ma2like=function(theta)
```

```
{
n=length(obs)
sigma=toeplitz(c(1+theta[1]^2+theta[2]^2
,theta[1]+theta[1]*theta[2],theta[2],rep(0,n-3)))
if (theta[2]<1 & theta[2]>-1 & theta[1]>-theta[2]-1 & theta[1]<theta[2]+1)
dmnorm(obs,rep(0,n),sigma) else 0
}
```

```
len=100
A=seq(-2,2,length=len)
B=seq(-1,1,length=len)
Z=expand.grid(A,B)
Z=as.matrix(Z)
resiobs=rep(0,len*len)
for (i in 1:(len*len)) resiobs[i]=ma2like(Z[i,])

cil=c("steelblue","tomato","gold")
par(mfrow=c(1,2))
li1=quantile(dede1,c(0.01,0.001,0.0001))
plot(0,0,cex=.4,pch=19,col="wheat"
,xlab=expression(theta[1]),ylab=expression(theta[2]),
xlim=c(-2,2),ylim=c(-1,1))
polygon(c(-2,0,2,-2),c(1,-1,1,1),col="wheat",border=F)
for (l in 1:3) points(theta[dede1<=li1[l],],col=cil[l],pch=19,cex=.4)
contour(A,B,matrix(resiobs,len,len),add=TRUE,nlev=5,drawlabels=FALSE)
points(0.6,0.2,pch=19,cex=0.5,col="red")
li2=quantile(dede2,c(0.01,0.001,0.0001))
plot(0,0,cex=.4,pch=19,col="wheat"
,xlab=expression(theta[1]),ylab=expression(theta[2]),
xlim=c(-2,2),ylim=c(-1,1))
polygon(c(-2,0,2,-2),c(1,-1,1,1),col="wheat",border=F)
for (l in 1:3) points(theta[dede2<=li2[l],],col=cil[l],pch=19,cex=.4)
contour(A,B,matrix(resiobs,len,len),add=TRUE,nlev=4,drawlabels=FALSE)
points(0.6,0.2,pch=19,cex=0.5,col="red")
```

• دستور زیر مربوط به شکل ۴.۴ می باشد.

```

len=200
A=seq(-2,2,length=len)
resin1=rep(0,len)
for (i in 1:len)
{
cucu=function(x)
{
ni=length(x)
mi=ma2like(c(A[i],x[1]))
if (ni>1) for (j in 2:ni) mi=c(mi,ma2like(c(A[i],x[j])))
mi
}
resin1[i]=integrate(cucu,-1,1)$value
print(i)
}
mp1=approxfun(A,resin1)
c1=integrate(mp1,-2,2)$value
mp1=approxfun(A,resin1/c1)

B=seq(-1,1,length=len)
resin2=rep(0,len)
for (i in 1:len)
{
cucu=function(x)
{
ni=length(x)
mi=ma2like(c(x[1],B[i]))
if (ni>1) for (j in 2:ni) mi=c(mi,ma2like(c(x[j],B[i])))
mi
}
resin2[i]=integrate(cucu,-2,2)$value
print(i)
}

```

```

mp2=approxfun(B,resin2/c1)

par(mfrow=c(1,2))
li1=quantile(dede1,c(0.1,0.01,0.001))
curve(mp1,-2,2,lwd=1,ylab="",xlab=expression(theta[1]))
llty=c(1,3,5,1)
for (l in 1:3)lines(density(theta[dede1<=li1[l],1]
,bw=0.05),lty=llty[l],lwd=2)
legend("topleft",c(".1",".001",".01"),lwd=2,lty=c(1,3,5))
curve(mp2,-1,1,lwd=1,ylab="",xlab=expression(theta[2]))
for (l in 1:3) lines(density(theta[dede1<=li1[l],2]
,bw=0.057),lty=llty[l],lwd=2)
legend("topleft",c("","",""),lwd=2,lty=c(1,3,5))
#col=cil[l]

```

مثال سوم فصل چهار

- دستورات زیر مربوط به شکل ۶.۴ برای $N = 10^5$ داده شبیه‌سازی شده است.

```

require(smfsb)
data(LVdata)
N=1e5
message(paste("N =",N))
prior=cbind(th1=exp(runif(N,-6,2)),
th2=exp(runif(N,-6,2)),th3=exp(runif(N,-6,2)))
rows=lapply(1:N,function(i){prior[i,]})
message("starting simulation")
samples=lapply(rows,function(th)
{simTs(c(50,100),0,30,2,stepLVc,th)})
message("finished simulation")
distance<-function(ts)
{
diff=ts-LVperfect
sum(diff*diff)
}
message("computing distances")

```



```

dist=lapply(samples,distance)
message("distances computed")
dist=sapply(dist,c)
cutoff=quantile(dist,1000/N)
post=prior[dist<cutoff,]
op=par(mfrow=c(2,3))
apply(post,2,hist,30)
apply(log(post),2,hist,30)
par(op)
#####
system.time
user      system    elapsed
489.80    0.71      503.31

```

- دستورات زیر مربوط به شکل ۷.۴ و با استفاده از بسته parallel در R است.

```

require(smfsb)
require(parallel)
options(mc.cores=4)
data(LVdata)
N=1e5
bs=1e4
batches=N/bs
message(paste("N =",N," | bs =",bs," | batches =",batches))
distance<-function(ts)
{
diff=ts[,1]-LVprey
sum(diff*diff)
}
post=NULL
for (i in 1:batches) {
message(paste("batch",i,"of",batches))
prior=cbind(th1=exp(runif(bs,-6,2)),
th2=exp(runif(bs,-6,2)),th3=exp(runif(bs,-6,2)))
rows=lapply(1:bs,function(i){prior[i,]})

```

```

samples=mclapply(rows,function(th)
{simTs(c(50,100),0,30,2,stepLVc,th)})
dist=mclapply(samples,distance)
dist=sapply(dist,c)
cutoff=quantile(dist,1000/N)
post=rbind(post,prior[dist<cutoff,])
}
message(paste("Finished. Kept",
dim(post)[1],"simulations"))
op=par(mfrow=c(2,3))
apply(post,2,hist,30)
apply(log(post),2,hist,30)
par(op)
#####
system.time
user      system    elapsed
463.37    0.11      465.65

```

- دستورات زیر مربوط به شکل ۸.۴ برای $N = 10^7$ داده شبیه‌سازی شده است.

```

require(smfsb)
require(parallel)
options(mc.cores=4)
data(LVdata)
N=1e7
bs=1e5
batches=N/bs
message(paste("N =",N," | bs =",bs," | batches =",batches))
ssinit <- function(vec)
{
ac23=as.vector(acf(vec,lag.max=2,plot=FALSE)$acf)[2:3]
c(mean(vec),log(var(vec)+1),ac23)
}
ssi <- function(ts)
{

```

```

c(ssinit(ts[,1]),ssinit(ts[,2]),cor(ts[,1],ts[,2]))
}
message("Batch 0: Pilot run batch")
prior=cbind(th1=exp(runif(bs,-6,2)),
th2=exp(runif(bs,-6,2)),th3=exp(runif(bs,-6,2)))
rows=lapply(1:bs,function(i){prior[i,]})
samples=mclapply(rows,function(th)
{simTs(c(50,100),0,30,2,stepLVc,th)})
sumstats=mclapply(samples,ssi)
sds=apply(sapply(sumstats,c),1,sd)
print(sds)
# now define a standardised distance
ss<-function(ts)
{
ssi(ts)/sds
}
ss0=ss(LVperfect)
distance <- function(ts)
{
diff=ss(ts)-ss0
sum(diff*diff)
}
post=NULL
for (i in 1:batches) {
message(paste("batch",i,"of",batches))
prior=cbind(th1=exp(runif(bs,-6,2)),
th2=exp(runif(bs,-6,2)),th3=exp(runif(bs,-6,2)))
rows=lapply(1:bs,function(i){prior[i,]})
samples=mclapply(rows,function(th)
{simTs(c(50,100),0,30,2,stepLVc,th)})
dist=mclapply(samples,distance)
dist=sapply(dist,c)
cutoff=quantile(dist,1000/N,na.rm=TRUE)
post=rbind(post,prior[dist<cutoff,])
}

```

```
}  
message(paste("Finished. Kept",dim(post)[1],"simulations"))  
th=c(th1 = 1, th2 = 0.005, th3 = 0.6)  
op=par(mfrow=c(2,3))  
for (i in 1:3) {  
  hist(post[,i],30,col=5,  
  main=paste("Posterior for theta[",i,"]",sep=""))  
  abline(v=th[i],lwd=2,col=2)  
}  
for (i in 1:3) {  
  hist(log(post[,i]),30,col=5,  
  main=paste("Posterior for log(theta[",i,"])",sep=""))  
  abline(v=log(th[i]),lwd=2,col=2)  
}  
par(op)  
#####  
system.time  
user      system    elapsed  
38692.07  11.28     70334.63
```

مراجع

- [۱] بارتل جی.آر، (۱۳۵۴)، ”اصول آنالیز حقیقی“، زعفرانی ج.، چاپ دوم، انتشارات دانشگاه تهران.
- [2] Allingham, D., King, R.A.R. and Mengersen, K.L. (2009). Bayesian estimation of quantile distributions, *Statistics and Computing*. 19(2), pp: 189-201.
- [3] Asmussen, Søren; Glynn, Peter, W. (2007). *Stochastic Simulation: Algorithms and Analysis*, *Stochastic Modelling and Applied Probability*. 57, pp: 242-258.
- [4] Baber, S., Voss, J., Webster, M. (2014). The Rate of Convergence for Approximate Bayesian Computation, *Euclid Mathematics and Statistics*. 9, pp: 80-105.
- [5] Bandyopadhyay, S.P., Forster, R.M. (2011). *Philosophy of Statistics*, New Hampshire. 7, pp: 354-361, London.
- [6] Beaumont, M., Zhang, W., Balding, D. (2002). Approximate Bayesian computation in population genetics., *Genetics*. 162(4), pp: 2025-2035.
- [7] Beaumont, M., Cornuet, J.M., Marin, J.M., Robert, C. (2009). Adaptive approximate Bayesian computation, *Biometrika*. 96(4), pp: 983-990.
- [8] Beaumont, M., Nielsen, R., Robert, C., Hey, J., Gaggiotti, O., Knowles, L., Estoup, A., Mahesh, P., Coranders, J., Hickerson, M., Sisson, S., Fagundes, N., Chikhi, L., Beerli, P., Vitalis, R., Cornuet, J.-M., Huelsenbeck, J., Foll, M., Yang, Z., Rousset, F., Balding, D., Excoffier, L. (2010). In defense of model-based inference in phylogeography, *Molecular and Ecology*. 19(3), pp: 436-446.
- [9] Berger, J.O., Wolpert, R.L. (1988). *The Likelihood Principle* (2nd ed.), Haywood, CA: The Institute of Mathematical Statistics, New York.
- [10] Blum, M. (2010). Approximate Bayesian computation: A non-parametric perspective, *American Statistical Association*. 105(491), pp: 1178-1187.
- [11] Boys, J.R., Wilkinson, J.D., Kirkwood, L.B.T. (2008). Bayesian inference for a discretely observed stochastic kinetic model, *Statistics and Computing*. 18, pp: 125-135.
- [12] Brown, W.A. (1991). *Matrices and Vector Spaces*, M. Dekker, New York.

- [13] Bu, Z., Callaway, E.J.D. (2011). Proteins move: Protein dynamics and long-range allostery in cell signaling, *Advances in Protein Chemistry and Structural Biology*. 83, pp: 163-221.
- [14] Casella, G, George, I.E. (1992). Explaining the Gibbs sampler, *The American Statistician*. 46, pp: 167-174.
- [15] Choi, C., Hall, W.J. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models, *Annals of Statistics*. 24(2), pp: 841-861.
- [16] Cisewski, J. (2014). *Approximate Bayesian Computation in Astronomy*. Carnegie Mellon University. 20(2).
- [17] Cornuet, J.M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.M., Balding, D.J., Guillemaud, T., Estoup, A. (2008). Inferring population history with DIYABC: a user-friendly approach to approximate Bayesian computation, *Bioinformatics*. 24(23), pp: 2713–2719.
- [18] Csilléry, K., Blum, M., Gaggiotti, O., François, O. (2010a). Approximate Bayesian computation in practice. *Trends in Ecology and Evolution*. 25(7), pp: 410–418.
- [19] Dean, T.A., Singh, S.S., Jasra, A., Peters, G.W. (2011). Parameter estimation for hidden Markov models with intractable likelihoods. 41(4), pp: 970-987.
- [20] Diggle, P. J. and Gratton, R.J. (1984). Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society. Series B*, 46(2), pp: 193-227.
- [21] Duren, P. (1970), *Theory of Hp-Spaces*, Academic Press, New York.
- [22] Fearnhead, P., Prangle, D. (2010). Semi-automatic approximate Bayesian computation, *Journal of the Royal Statistical Society, Series B*. 74(3), pp: 419-474.
- [23] Fisher RA. (1926). The arrangement of field experiments, *Journal MINISTRY OF AGRICULTURE G.B.* 33, pp: 700-725.
- [24] Gelfand, A., Smith, A. (1990). Sampling based approaches to calculating marginal densities, *Journals American Statistical Association*. 85(410), pp: 398-409.
- [25] Gelman, A. and Meng, L.X. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling, *Statistical Science*. 13(2), pp: 163-185.
- [26] Geman, S., Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 6, pp: 721–741.

- [27] Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions, *The Journal of Physical Chemistry*. 81, pp: 2340–2361.
- [28] Green, P.J. (1995). Reversible-jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*. 82(4), pp: 711-732. doi:10.1093/biomet/82.4.711.
- [29] Grelaud, A., Marin, J.M., Robert, C., Rodolphe, F., Tally, F. (2009). Likelihood-freemethods formodel choice in Gibbs random fields, *Bayesian Anal.* 3(2), pp: 427–442.
- [30] Hannig, J., Iyer, H., Patterson, P. (2006). Fiducial Generalized Confidence Intervals, *Journal of the American Statistical*. 101, 254-269.
- [31] Haynes, M. (1998). Flexible distributions and statistical models in ranking and selection procedures, with applications. PhD thesis, Queensland University of Technology.
- [32] Jaakkola, T., Jordan, M. (2000). Bayesian parameter estimation via variational methods, *Statistics and Computing*. 10(1), 25–37.
- [33] Joyce, P., Marjoram, P. (2008). Approximately sufficient statistics and Bayesian computation, *Statistical Applications in Genetics and Molecular Biology*. 7(1), pp: 1544-1565 . DOI: 10.2202/1544-6115.1389.
- [34] Kinderman, J.A., Monahan, F.J., Ramage, G.J. (1977). Computer methods for sampling from Students distribution, *Mathematic of Computation*. 31(140), pp: 1009-1018.
- [35] Leuenberger, C., Wegmann, D., Excoffier, L. (2010). Bayesian computation and model selection in population genetics, *Genetics*. 184(1), pp: 243–252.
- [36] Levy, P., (1924). Theorie des erreurs. La loi de Gauss et les lois exceptionnelles, *Bulletin de la Société Mathématique de France* 52, 49-85.
- [37] Li, Y., and Ghosh, K.S. (2013). Efficient Sampling Methods for Truncated Multivariate Normal and Student-t Distributions Subject to Linear Inequality Constraints, NC State Department of Statistics, Technical Reports. 27(1), pp: 601-612.
- [38] Liu, S.J. (1996). Metropolisised independent sampling with comparisons to rejection sampling and importance sampling, *Statistics and Computing*. pp: 1-35.
- [39] Marin, J.M., Robert, C. (2007). *Bayesian Core*. Springer, New York.
- [40] Marin, J.M., Robert, C. (2010). Importance sampling methods for Bayesian discrimination between embedded models, In: Chen, M.-H., Dey, D., Müller, P., Sun, D., Ye, K. (eds.) *Frontiers of Statistical Decision Making and Bayesian Analysis*, pp: 513-527.

- [41] Marin, J.M., Pudlo, P., Robert C., Ryder, R. (2012). Approximate Bayesian Computation Methods, *Statistics and Computing*. 21, 289-291.
- [42] Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods, *Proceedings of the National Academy of Sciences*. 100(26), 15324–15328.
- [43] Martin, M.G., McCabe, B., Maneesoonthorn, W., Robert, C. (2014). Approximate Bayesian Computation in State Space Models. *math.ST*. arXiv:1409.8363
- [44] Murray, I., Ghahramani, Z., MacKay, D. (2006). MCMC for doubly-intractable distributions, In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, Arlington, Virginia, AUAI Press (2006). pp: 359-366.
- [45] Neyman, Jerzy; Pearson, Egon S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 231, pp: 694-706. doi:10.1098/rsta.1933.0009.
- [46] O'Hagan, A. (1987). Monte Carlo is fundamentally unsound, *The Statistician*, 36, pp: 247-249.
- [47] Pesarin, F. (2001). *Multivariate Permutation Tests: With Applications In Biostatistics*, Wiley, New York.
- [48] Peters, G. (2009). *Advances in Approximate Bayesian Computation and trans-dimensional sampling methodology*. PhD thesis, University of NSW.
- [49] Peters, W.G., Sisson, A.S., Fan, Y. (2012). Likelihood-free Bayesian inference for -stable models, *Computational Statistics and Data Analysis*. 56, pp: 3743- 3756.
- [50] Prangle, D. (2011). *Summary Statistics and Sequential Methods for Approximate Bayesian Computation*. PhD thesis, University of Lancaster.
- [51] Pritchard, J., Seielstad, M., Perez-Lezaun, A., Feldman, M. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites, *Molecular Biology and Evolution*. 16(12), pp: 1791–1798.
- [52] Ratmann, O. (2009). *ABC under model uncertainty*. PhD thesis, Imperial College, London.
- [53] Rayner, G.D. and MacGillivray, H.L. (2002). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions, *Statistics and Computing*. 12(1), pp: 57–75.
- [54] Richey, M. (2010). The Evolution of Markov Chain Monte Carlo Methods, *The American Mathematical Monthly*, Mathematical Association of America. 117(5), pp: 383-413. doi:10.4169/000298910x485923.

- [55] Robert, C. (2001). *The Bayesian Choice*, 2nd edn. Springer, New York.
- [56] Robert, C., Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edn. Springer, New York.
- [57] Robert, C.P., Mengersen, K., Chen, C. (2010). Model choice versus model criticism, *Proceedings of the National Academy of Sciences*. 107(3), E5 (2010).
- [58] Robert, C.P., Cornuet, J.M., Marin, J.M., Pillai, N. (2011). Lack of confidence in ABC model choice, *Proceedings of the National Academy of Sciences*. 108(37), pp: 15112-15117.
- [59] Rubin, D. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician, *Annals of Statistics*. 12(4), pp: 1151-1172.
- [60] Rudin, W. (1987). *Real and complex analysis*. International Series in Pure and Applied Mathematics (3rd ed.), McGrawHill, London.
- [61] Rue, H., Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Monographs on Statistics and Applied Probability. 104, Chapman and Hall, London.
- [62] Samorodnitsky, G., Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall/CRC, London.
- [63] Sham, P.C., Purcell, M.S. (2014). Statistical power and significance testing in large-scale genetic studies, *Nature Reviews Genetics*. 15 (5), pp: 335-346. doi:10.1038/nrg3706.
- [64] Sigman, Karl. (2007). *Acceptance-Rejection Method*, Columbia University, New York.
- [65] Sisson, S.A., Fan, Y., Tanaka, M. (2007). Sequential Monte Carlo without likelihoods, *Proceedings of the National Academy of Sciences*. 104(6), pp: 1760–1765.
- [66] Sisson, S.A., Fan, Y., Tanaka, M. (2009). Sequential Monte Carlo without likelihoods: Errata, *Proceedings of the National Academy of Sciences* 106(39). 1760-1765.
- [67] Smith, R.L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions, *Operations Research*. 32, pp: 1296–1308. doi:10.1287/opre.32.6.1296.
- [68] Stein, CM. (1981). Estimation of the mean of a multivariate normal distribution, the *annals of Statistics*. 9(6), pp:1135-1151.
- [69] Stewart, S., Gomez, A.W., Armstrong, B.E., Henner, A., and Stankunas, K. (2014). Sequential and opposing activities of Wnt and BMP coordinate zebrafish bone regeneration, *Zebrafish Model Organism Database*, 6(3), pp :482-498.

-
- [70] Tavaré, S., Balding, D., Griffith, R., Donnelly, P. (1997). Inferring coalescence times from DNA sequence data, *Genetics*. 145(2), pp: 505-518.
- [71] Tierney, L., Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities, *Journals - American Statistical Association*. 81(393), pp: 82–86.
- [72] Wald, A. (1941). Asymptotically most powerful tests of statistical hypotheses, *The Annals of Mathematical Statistics*. London.
- [73] Walters, Peter. (1982). An introduction to ergodic theory, *Graduate Texts in Mathematics*, 79, Springer-Verlag, USA.
- [74] Wilkinson, D. (2006). *Stochastic modelling for systems biology*. Chapman and Hall/CRC. London.
- [75] Wilkinson, R.D. (2008). Approximate Bayesian computation gives exact results under the assumption of model error, *Statistical Applications in Genetics and Molecular Biology*. 12(2), pp: 29-41.
- [76] Vrugt, A.J., and Sadegh, M. (2013). Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, *Water Resources Research*. 49, pp: 4335-4345.

واژه‌نامه فارسی به انگلیسی

Instrumental	ابزاری
Paradims	الگو
Pilot design	آزمایشی اولیه
Permutation test	آزمون جایگشتی
Significance test	آزمون معناداری
Cross-validation	اعتبار سنجی
Summary statistic	آماره خلاصه شده
Entropy	آنتروپی
Median absolute deviation	انحراف میانه مطلق
Reparametrization	بازپارامتری کردن
Recurrent	بازگشتی
UMVUE	برآوردگر به‌طور یکنواخت ناریب با کمترین واریانس
Maximum Likelihood Estimation	برآوردیابی درست‌نمایی ماکزیمم
Tuning parameter	پارامتر تنظیم‌کننده
Envelop	پوش
Bandwidth	پهنار نوار
Noninformative prior	پیشین ناآگاهی‌بخش
Loss function	تابع زیان
Quadratic loos function	تابع زیان توان دوم
Risk function	تابع مخاطره
Tail decay	تباهی دم
Analysis of variance	تحلیل واریانس
Irreducible	تحویل ناپذیری
Adaptive Gauss Hermit Qudrature	تربیع‌بندی گاوس هرमित سازوار
Randomization	تصادفی‌سازی

Inadmissible decision rule	تصمیم ناپذیرفتنی
Proposal distribution	توزیع پیشنهادی
Quantile distribution	توزیع چندکی
Normalization constant	ثابت نرمال‌ساز
Phytogeography	جغرافیای گیاهی
Prior predictive density	چگالی پیشگوی پیشین
Spectral density	چگالی طیفی
Skewness	چولگی
Effective Sample Size	حجم نمونه موثر
Trade-off	داد و ستد
Burn-in	داغیدن
Free likelihood	درست‌نمای آزاد
Empirical likelihood	درست‌نمایی تجربی
Marginal likelihood	درست‌نمای حاشیه‌ای
Periodogram	دوره‌نگار
Intractable	رام‌نشدنی
Full rank	رتبه کامل
Allowed classes	رده‌های مجاز
Local linear regression	رگرسیون خطی موضعی
Subadditivity	زیرجمع‌ی
Biology	زیست‌شناسی
system biology	سیستم زیست‌شناسی
Detailed balance conditon	شرط تعادل دقیق
predator-pray	شکارچی-شکار
Evidence	شهود
Asymmetry	عدم تقارن
nonidentifiable	غیرقابل تشخیص
Fiducial intervals	فواصل اتکایی
Deductive	قیاسی
Calibration	کالیبدن
Kurtosis	کشیدگی
Lebesgue differentiation theorem	قضیه مشتق‌پذیری لبگ

Partial least squar.....	کمترین توان دوم جزیی
Approximate Bayesian Computational.....	محاسبات بیزی تقریبی
Grid.....	مشبکه
Posetive definite.....	معین مثبت
Scalability absolute.....	مقیاس پذیری مطلق
Markov Chain Monte Carlo.....	مونت کارلوی زنجیر مارکوف
Mean squared error.....	میانگین توان دوم خطا
Jensen's inequality.....	نامساوی جنسن
Inequality triangle.....	نامساوی مثلث
Yonug inequilty.....	نامساوی یونگ
Accept-Reject Sampling.....	نمونه‌گیری پذیرش-رد
Importance Sampling.....	نمونه‌گیری نقاط مهم
Noisy.....	نوفه
Seminorm.....	نیم‌نرم
Semi automatic.....	نیمه خودکار
Imbalanced weight.....	وزن‌های نامتعادل
Kernel.....	هسته
Transition kernel.....	هسته انتقال
Homogeneity absolute.....	همگنی مطلق
Smoothing.....	هموار ساز
Epidemiology.....	همه‌گیرشناسی
Machine learning.....	یادگیری ماشینی

واژه‌نامه انگلیسی به فارسی

Absolute error loss	زیان قدرمطلق خطا
Accept-Reject Sampling	نمونه‌گیری پذیرش-رد
Adaptive Gauss Hermit Quadrature	تربیع‌بندی گاوس هرمیت سازوار
Allowed classes	رده‌های مجاز
Analysis of variance	تحلیل واریانس
Approximate Bayesian Computational	محاسبات بیزی تقریبی
Asymmetry	عدم تقارن
Bandwidth	پهنای نواری
Biology	زیست‌شناسی
Burn-in	داغیدن
Calibration	کالیبره کردن
Cross-validation	اعتبارسنجی
Deductive	قیاسی
Detailed balance condition	شرط تعادل دقیق
Effective Sample Size	حجم نمونه موثر
Empirical likelihood	درست‌نمایی تجربی
Evidence	شهود
Entropy	آنتروپی
Envelop	پوش
Epidemiology	همه‌گیرشناسی
Fiducial intervals	فواصل اتکایی
Free likelihood	درست‌نمای آزاد
Full rank	رتبه کامل
function hazard	تابع مخاطره
Grid	مشبکه

Homogeneity absolute	همگنی مطلق
Imbalanced weight	وزن‌های نامتعادل
Importance Sampling	نمونه‌گیری نقاط مهم
Inadmissible decision rule	تصمیم ناپذیرفتنی
Inequality triangle	نامساوی مثلث
Instrumental	ابزاری
Intractable	رام‌نشدنی
Irreducible	تحویل ناپذیری
Jensen's inequality	نامساوی جنسن
Kurtosis	کشیدگی
Lebesgue differentiation theorem	قضیه مشتق‌پذیری لبگ
Local linear regression	رگرسیون خطی موضعی
Loss function	تابع زیان
Machine learning	یادگیری ماشینی
Marginal likelihood	درستمای حاشیه‌ای
Markov Chain Monte Carlo	مونت کارلوی زنجیر مارکوف
Maximum Likelihood Estimation	برآوردیابی درستمایی ماکزیمم
Mean squared error	میانگین توان دوم خطا
Median absolute deviation	انحراف میانه مطلق
Noisy	نوفه
Noninformative prior	پیشین ناآگاهی‌بخش
Normalization constant	ثابت نرمال‌ساز
Paradims	الگو
Partial least squar	کمترین توان دوم جزئی
Periodogram	دوره‌نگار
Permutation test	آزمون جایگشتی
Phytogeography	جغرافیای گیاهی
Pilot design	آزمایشی اولیه
Posetive definite	معین مثبت
predator-pray	شکارچی-شکار
Prior predictive density	چگالی پیشگوی پیشین
Proposal distribution	توزیع پیشنهادی

Quantile distribution.....	توزیع چندکی
Quadratic loss function.....	تابع زیان توان دوم
Randomization.....	تصادفی سازی
Recurrent.....	بازگشتی
Reparametrization.....	بازپارامتری کردن
Risk function.....	تابع مخاطره
Semi automatic.....	نیمه خودکار
Scalability absolute.....	مقیاس پذیری مطلق
Seminorm.....	نیم نرم
Significance test.....	آزمون معناداری
system biology.....	سیستم زیست شناسی
Skewness.....	چولگی
Smoothing.....	هموار ساز
Spectral density.....	چگالی طیفی
Subadditivity.....	زیر جمعیتی
Summary statistic.....	آماره خلاصه شده
Tail decay.....	تباهی دم
Trade-off.....	داد و ستد
Transition kernel.....	هسته انتقال
Tuning parameter.....	پارامتر تنظیم کننده
UMVUE.....	برآوردگر به طور یکنواخت ناریب با کمترین واریانس
Yonug inequality.....	نامساوی یونگ

Aabstract

Approximate Bayesian Computation (ABC) methods, also known as likelihood-free sampling techniques, have been appeared in the past ten years as the most satisfactory approach to Bayesian analysis of complex and intractable likelihood problems. ABC was used first in genetics, then in a broader spectrum of applications. Over the past few years, both practical and theoretical issues of these methods, in statistics and other disciplines, are desired. Due to the massive and unprecedented reception of such methods, in this thesis, we present different algorithms and review some theoretical aspects for ABC methods. Then, we show some examples of their applications in real and simulated situations.

Keywords Likelihood-free methods, Bayesian statistics, ABC methodology, Noisy ABC.



Shahrood University
Faculty Of Mathematical Sciences

Dissertation Submitted in Partial
Fulfillment of The Requirements For The
Degree of Master of Science in
Statistics

Approximate Bayesian Computation: Methods and Applications

Supervisor

Dr.Hossein Baghishani

Advisor

Dr.Negar Eghbal

by

Saeed Afchangi

February 2015