

انقلاب بی‌زی در علم ژنتیک

THE BAYESIAN REVOLUTION IN GENETICS

Nature Reviews Genetics, 2004

مترجم: صادق علیجانی

تبریز، دانشگاه تبریز، دانشکده کشاورزی، گروه علوم دامی

پست الکترونیکی: sad-ali@tabrizu.ac.ir

چکیده

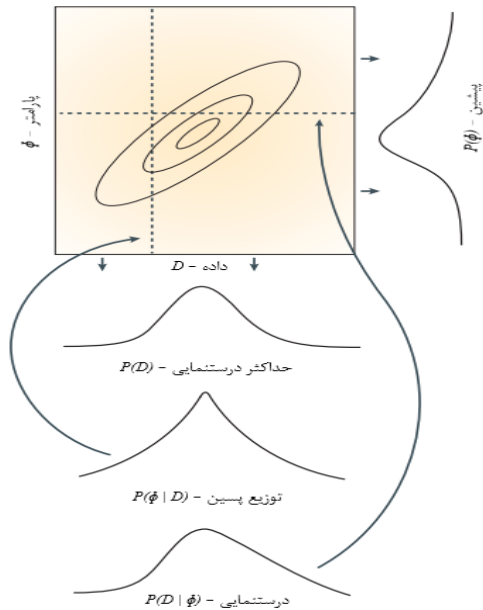
بیشتر متخصصین ژنتیک از نظر فکری از تفکر آمار کلاسیک (فراوانی‌گرایی یا مرسوم) برخوردار هستند. این نوع تفکر آماری شامل آزمون فرض، برآورد و تشکیل فواصل اطمینان برای پارامترهای ناشناخته است. در گذشته این چارچوب فکری منجر به پیشرفتهای زیادی در زمینه‌های مختلف شده است. اما تفکر آماری دیگر یعنی آمار بی‌زی چه نوع آورده‌های جدیدی برای ژنتیک دانان می‌تواند داشته باشد؟ آمار بی‌زی این امکان را برای دانشمندان ایجاد می‌کند که به آسانی دانش پیشین خود را در تجزیه و تحلیل آماری داده‌ها دخالت دهند. با این حال بالا بودن توان محاسباتی مورد نیاز این نوع تجزیه و تحلیل‌ها، استفاده از این روش را در علم ژنتیک محدود کرده بود. هم‌اکنون محدودیت‌های محاسباتی به مقدار زیادی برطرف شده و مزایای اساسی روش‌های بی‌زی باعث شده تا این روش‌ها در تعداد فزاینده‌ای از حوزه‌ها به عنوان روش پیشنهادی در تجزیه و تحلیل داده‌های ژنتیکی مورد استفاده قرار گیرند.

که خود یک بستر در حال رشد سریع است، تکمیل می‌شود و ژنومیک عملکردی نیز دارای خیلی از خصائص (و مسائل) دیگر علوم مشاهده‌ای است که نیازمند تجزیه و تحلیل‌های احتمالی مشابهی می‌باشند. معمولاً داده‌های ژنتیکی ماحصل فرایندهای پیچیده با مکانیسم‌های متعددی هستند که می‌توانند داده‌های مشاهده شده را تولید نمایند، بنابراین این سؤال مطرح می‌شود که بهترین راه انتخاب عوامل مسبب ممکن کدامند؟ به عنوان مثال استفاده از داده‌های ژنتیکی برای تعیین ساختار مخفی جمعیتها (یعنی افراد با اجداد جمعیتی مختلف، مثلاً مناطق مختلف جغرافیایی که از آنها منشاء گرفته‌اند) در نظر بگیرید. محاسبه شانس اینکه فردی با ژنوتیپ خاص در جمعیتی به دنیا آمده که متفاوت از جمعیتی است که از آن نمونه‌گیری شده است (یعنی اینکه یک مهاجر است) در میان خیلی از چیزها به فراوانی ژنی آن جمعیت بستگی دارد. استنباط در مورد فراوانی‌های ژنی جمعیت بنوبه خود بستگی به استنباط در مورد جمعیت‌هایی دارد که دیگر افراد نمونه استخراج شده، از آنها سر منشاء گرفته‌اند (با توجه به ژنوتیپ‌شان) که این نیز بنوبه خود بستگی به فراوانی‌های ژنی استنباط شده برای تمامی دیگر جمعیت‌ها دارد و الی آخر. برای حل این نوع از مسائل (یعنی مدل‌هایی که در

در چندین شاخه از علم ژنتیک، همانند دیگر حوزه‌های زیست‌شناسی، فرایندهای پیچیده زیادی داده‌ها را تحت تاثیر قرار می‌دهند. علم ژنتیک نظریه‌های ریاضی توانمندی را برای بررسی این پیچیدگی ایجاد نموده است. در اغلب موارد این امکان وجود دارد که با استفاده از این ابزارهای نظری مدل‌های واقعی را تشکیل داده و بتوان داده‌ها را به صورت تابعی از این فرایندها تبیین کرد. غالباً اولین قدم در مطالعه فرایندهای دخالت‌کننده و مهیا نمودن پایه استنباط آماری، فرموله کردن چنین مدلی است. اکثر خصوصیات ژنتیکی افراد، جمعیت‌ها یا گونه‌ها (همانند ژنوتیپ افراد، فراوانی‌های ژنی جمعیت و چندشکلی‌های توالی DNA) نتیجه فرایندهایی هستند که ذاتاً تصادفی هستند و بنابراین نمی‌توانند بدون استفاده از مدل‌های احتمالی مورد بررسی قرار گیرند. البته این به بدین معنی نیست که هر جنبه از زیست‌شناسی مولکولی باید با استفاده از مدل‌های احتمالی مورد مطالعه قرار گیرد. به عنوان مثال در سطح بیوشیمیایی، مسیرهای ویژه‌ی بیان ژن می‌توانند تحت شرایط کم و بیش کنترل شده (حداقل برای تعدادی از محققین کاربردی) که نیازمند هیچگونه تجزیه و تحلیل آماری نیستند، مورد مطالعه قرار گیرند. با این حال حتی چنین مطالعات آزمایشی نیز به صورت فزاینده‌ای با ژنومیک عملکردی^۱

^۱ functional genomics

طرق مختلف و برای استنباط در مورد پارامترهای مدل احتمالی است (شکل ۱). هدف اصلی استنباط بیزی محاسبه توزیع پسین پارامترها یعنی منحنی توزیع‌های شرطی پارامترها (به شرط داده‌ها) است.



شکل ۱- خصوصیات اساسی مورد تاکید استنباط بیزی. فرض می‌شود داده‌های D می‌توانند هر مقداری را در طول محور اندازه‌گیری به خود بگیرند. همین‌طور، مقدار پارامتر ϕ نیز می‌تواند هر مقداری را در طول محور γ به خود بگیرد. استنباط بیزی شامل ایجاد توزیع توأم پارامترها و داده‌ها، $P(D, \phi)$ ، به شکل فواصل هم‌تراز در شکل است. این توزیع توأم می‌تواند به سادگی از طریق ضرب کردن توزیع پیشین $P(\phi)$ و درست‌نمایی $P(D|\phi)$ به دست آید. نوعاً، درست‌نمایی برخاسته از مدل آماری است که برحسب ضرورت برای تعیین چگونگی تعیین داده‌ها به وسیله پارامترها، در نظر گرفته می‌شود. توزیع پیشین یک توزیع مفروض برای پارامتر است که با استفاده از دانش زمینه‌ای حاصل می‌شود. فلش‌ها در شکل نشان می‌دهند که توزیع‌های حاشیه‌ای از جمع زدن (انتگرال‌گیری) توزیع توأم بر روی داده‌ها، یعنی به دست آوردن توزیع پیشین (توزیع دست راست توزیع توأم) یا بر روی مقادیر پارامترها که منجر به تولید درست‌نمایی‌های حاشیه‌ای می‌شود، (اولین توزیع که مستقیماً در زیر توزیع توأم قرار دارد)، ایجاد می‌شوند. توزیع‌های شرطی که توسط علامت «|» نشان داده می‌شوند، توسط خطوط منقطع در شکل نمایش داده شده و یک برش از توزیع توأم و بازمقیاس‌بیدن آن توزیع به صورتی است که مجموع (انتگرال) مقادیر ممکن برابر با یک شود. عامل مقیاس مورد نیاز نیز خود توزیع حاشیه‌ای است. هر توزیع شرطی به سادگی برابر با تقسیم توزیع توأم بر توزیع حاشیه‌ای می‌باشد. به عنوان مثال، درست‌نمایی می‌تواند از طریق تقسیم توزیع توأم بر توزیع پیشین به دست آید. توزیع پسین، $P(\phi|D)$ ، کمیتی که ما در استنباط بیزی به آن نیاز داریم، برابر با توزیع توأم تقسیم شده بر درست‌نمایی حاشیه‌ای می‌باشد. محاسبه درست

آنها چندین پارامتر وابسته به یکدیگر دخیل هستند) استنباط بیزی روش مناسبی است.

در این مرور کاربرد روش بیزی در تجزیه و تحلیل‌های ژنتیکی با دیگر روش‌هایی که در آنها دیگر چارچوب‌های آماری مورد استفاده قرار گرفته است، مقایسه می‌شود. تلاش خواهیم کرد تا توضیح دهیم چرا استفاده از روش‌های بیزی در تعدادی از شاخه‌های علوم، در طی دهه گذشته، افزایش پیدا کرده و برخی جنبه‌های مسائل ژنتیکی را که باعث شده است تا استدلال بیزی به صورت ویژه‌ای جذاب شود مشخص خواهیم کرد. جذابیت بالقوه روش‌های بیزی عبارت از توانایی این روش در دخیل کردن اطلاعات زمینه‌ای در تعیین مدل است. با این حال استدلال ما این است که معروفیت اخیر روش‌های بیزی عمدتاً واقعیت‌گرایی آن بوده و قابل توضیح است که چگونه می‌توان به وسیله آن و به آسانی با استفاده از فناوری‌های محاسباتی و کاربرد محاسبات گسترده زنجیره‌های مارکف منت کارلو^۱ (MCMC) از عهده مسائل درست‌نمایی پیچیده برآمد. برای نشان دادن این موضوع کاربردهای اخیر استنباط بیزی در سه حوزه جدید از تجزیه و تحلیل‌های ژنتیکی شامل: ژنتیک جمعیت، ژنومیکس و ژنتیک انسانی توضیح داده خواهد شد.

اصول استنباط بیزی

در ذات دیدگاه بیزی تفاوت منطقی بین پارامترهای مدل و داده‌ها وجود ندارد؛ هر دو، متغیرهای تصادفی هستند که توزیع توأم آنها توسط مدل احتمالی تعیین می‌شود. از این دیدگاه، داده‌ها، متغیرهای تصادفی مشاهده شده بوده، پارامترها متغیرهای تصادفی مشاهده نشده هستند. توزیع توأم نیز برابر با حاصل ضرب درست‌نمایی و توزیع پیشین می‌باشد. توزیع پیشین در برگیرنده اطلاعات مربوط به مقادیر یک پارامتر به شکل توزیع احتمال و قبل از بررسی داده‌ها است. درست‌نمایی، توزیع شرطی است که احتمال داده‌های مشاهده شده را به شرط هر مقدار خاص از پارامترها معین کرده و بر اساس مدل فرآیندهای زمینه‌ای ایجاد می‌شود. تابع‌های توزیع پیشین و درست‌نمایی، توأم با هم، تمامی اطلاعات موجود مربوط به پارامترها را دربرمی‌گیرند. آمار بیزی شامل استفاده از این توزیع توأم بر

¹ Markov Chain Monte Carlo

و انتخاب برآورد نقطه‌ای طوری انجام می‌شود که احتمال تولید داده‌ها به شرط پارامترها را بیشینه کند (یعنی، بیشینه سازی درست نمایی به صورت تابعی از پارامترها برای یک مجموعه ثابت از داده‌ها).

از نظر تاریخی استدلال‌های متعددی در مخالفت و حمایت از انواع مختلف استنباط‌ها وجود داشته است. اسناد قدیمی در ارتباط با روش بیزی حاکی از این است که نوعی عدم رضایت در انتخاب توزیع پیشین وجود دارد. با این حال، این مشکل در اصل تفاوتی با انتخاب تابع درست نمایی در روش حداکثر درست نمایی ندارد. در واقع همان‌طوری‌که در ادامه ثابت می‌شود، روش‌های بیزی جدید اغلب برای محاسبه احتمال‌های پسین به شرط داده‌ها از احتمال‌های پیشین صریح برای انواع توابع درست نمایی استفاده می‌کنند.

دلایل کاربردی استفاده از استنباط بیزی عبارتند از: الف) چنانچه مدل احتمال شامل چندین متغیر وابسته به هم، با مقادیر آنها محدود به دامنه خاصی از مقادیر، باشد (همان طور که در اغلب موارد در ژنتیک این چنین است)،

نمایی حاشیه‌ای (یعنی، انتگرال گیری‌هایی که توسط فلش‌های اشاره کننده به پایین از توزیع توأم نشان داده شده‌اند) نوعاً، مسئله ساز هستند.

برآورد نقطه‌ای (Point estimate) یک پارامتر با استفاده از برخی از خصوصیات توزیع پسین (معمولاً مد یا میانگین) به دست می‌آید. برآورد فاصله‌ای بیز برای یک پارامتر نیز می‌تواند با در نظر گرفتن یک مجموعه قابل اعتماد (Credible set) از مقادیر (یک مجموعه یا فاصله‌ای که در برگزیده پارامتر حقیقی با احتمال $1-\alpha$ بوده و α نیز سطح معنی داری از قبل تعیین شده، همانند $0/05$ می‌باشد). مثالی که در آن از استنباط بیزی برای منتسب نمودن فردی براساس ژنوتیپ خود فرد و از یک جمعیت با مرجع ناشناخته به جمعیتی که در آن به دنیا آمده است، در جدول ۱ نشان داده شده است.

دیگر روش‌های کاملاً شناخته شده غیر بیزی استنباط آماری، شامل روش حداکثر درست نمایی و روش گشتاورها است که پایه استنباط فراوانی‌گراها یا آمار کلاسیک به حساب می‌آید. در روش حداکثر درست نمایی، استنباط ما صرفاً براساس تابع درست نمایی انجام شده و اطلاعات پیشین هیچ‌گونه دخالتی در استنباط نداشته

جدول ۱- مثالی از استنباط بیزی: انتساب افراد به جمعیت‌ها

داده‌ها (متغیرهای مشاهده شده)							پارامترها (متغیرهای مشاهده نشده)
ژنوتیپ B			ژنوتیپ A				
احتمال پیشین	احتمال پسین	توزیع توأم	درست نمایی	احتمال پسین	توزیع توأم	درست نمایی	
	۰/۶۹		۰/۹۹	۰/۰۰۱۲		۰/۰۱	مهاجر
۰/۱		۰/۰۹۹			۰/۰۰۱		ساکن
	۰/۳۱		۰/۰۵	۰/۹۹۸۸		۰/۹۵	
۰/۹		۰/۰۴۵			۰/۸۵۵		
۱		۰/۱۲۴			۰/۸۵۶		احتمال داده‌ها

این مثال باید با توجه به شکل ۱ تفسیر شود. وضعیتی را در نظر می‌گیریم که در آن افراد هاپلوئید در جمعیتی قرار دارند که مهاجرت به آن با یک نرخ پایین صورت می‌گیرد. با توجه به اطلاعات زمینه‌ای، همانند داده‌هایی که از طریق نصب طوقه پا در پرندگان به دست آمده، چنین فکر می‌کنیم که احتمال اینکه هر فرد انتخاب شده به صورت تصادفی، ساکن باشد برابر $0/9$ و احتمال اینکه مهاجر باشد برابر $0/1$ است؛ این در واقع توزیع پیشین ماست (ستون آخر در سمت چپ). در این جمعیت در یک جایگاه ژنی دو آلل (A و B) وجود دارد. دوباره با توجه به اطلاعات زمینه‌ای، فکر می‌کنیم، در گروه مهاجر درست نمایی ژنوتیپ A برابر $0/01$ و در جمعیت ساکن برابر $0/95$ می‌باشد (اولین ستون تحت ژنوتیپ A). توزیع توأم نیز برابر با حاصل ضرب توزیع پیشین و درست نمایی (ستون‌های وسط تحت هر ژنوتیپ) است؛ این احتمال نشان دهنده احتمال یک مشاهده خاص می‌باشد. برای مثال، توزیع توأم یک مهاجر و با ژنوتیپ A برابر $0/01$ است. احتمال اینکه یک مشاهده از یک ژنوتیپ خاص باشد، صرفنظر از اینکه ساکن یا مهاجر باشد، برابر با حاشیه‌ی پایینی جدول است، این مقادیر حاشیه‌ای به وسیله جمع کردن توزیع‌های توأم برای تمام مقادیر پارامتری به دست می‌آید. به شرط مشاهده یک ژنوتیپ خاص، احتمال پسین اینکه مهاجر یا ساکن باشد (ستون‌های دست راستی قسمت هر ژنوتیپ) با محاسبه توزیع توأم و تغییر مقیاس آن از طریق تقسیم کردن توزیع توأم بر احتمال داده‌ها به صورتی که مجموع احتمال‌ها برابر با یک باشد، به دست می‌آید. بنابراین، چنانچه ما ژنوتیپ B را مشاهده کنیم، احتمال پسین این که آن فرد مهاجر باشد برابر $0/69$ خواهد بود (در حالی که قبل از این مشاهده برابر $0/1$ بود).

های آماری اولیه شامل محاسبه امیدریاضی برآوردگرهای گوناگون به صورت توابعی از پارامترهای مدل ژنتیکی و کاربرد روش گشتاورها بود. تا سالهای اخیر، روشهای درست نمایی برای حل مسائل ژنتیک جمعیت به کار برده نشده بود. توسعه نظریه «یکی کردن»^۴ بطور قوی چندین حوزه از علم ژنتیک جمعیت را تحت تأثیر قرار داد، این تئوری همانند روشهای اولیه، محاسبه امیدهای ریاضی را ممکن و در عین حال شبیه‌سازی سریع داده‌های نمونه‌ای را برای روش خودگردان سازی^۵ پارامتری مهیا نموده که این به نوبه خود امکان انجام محاسبات پیچیده برای تشکیل فواصل اطمینان و آزمون فرض روش فراوانی‌گراها را امکان‌پذیر می‌کند. اگرچه این تئوری در تمامی حوزه‌های تجزیه و تحلیل ژنتیکی جمعیت‌ها قابل استفاده نیست، اما ایجاد کننده پایه‌های اساسی برای محاسبات درست نمایی در مدل‌های دودمانی یا شجره‌ای بوده و استفاده از روش‌های بیزی برای استنباط در مورد تاریخچه جمعیتی، با استفاده از داده‌های ژنتیکی را امکان‌پذیر می‌کند. علاوه بر این، روش‌های بیزی در انتساب افراد به جمعیت‌هایی که از آنها منشأ گرفته‌اند و همین‌طور ردیابی اثر انتخاب بر ژن‌ها نیز مورد استفاده قرار می‌گیرد.

برآورد پارامترها در مدل‌های جمعیت شناسی. یکی از خصوصیات استنباط‌های ژنتیک جمعیتی این است که پارامترهای تابع درست نمایی همانند نرخ جهش (μ) و اندازه مؤثر جمعیت (N_e) تنها به صورت ضرب شده در هم (μN_e) در این تابع ظاهر می‌شوند، و این به این معنی است که این پارامترها غیر قابل تشخیص از هم هستند. در روش‌های استنباط غیر بیزی، چنانچه یکی از این پارامترها مورد بررسی قرار گیرند، عموماً برآورد نقطه‌ای «بهترین حدس» برای دیگر پارامتر مورد استفاده قرار گرفته و روش توانمندی برای دخالت دادن این عدم قطعیت وجود ندارد. توانمندی قابل استدلال روش بیزی این است که اطلاعات پیشین می‌توانند در به دست آوردن استنباط در مورد پارامترهای غیر قابل تشخیص، مورد استفاده قرار گیرند.

مدل‌های جمعیت شناسی اغلب دارای چندین پارامترند که از نظر مفهومی آسان‌تر است که استنباط در مورد آنها به صورت انفرادی و یا اینکه حداکثر به صورت توأم جفتی

در استنباط حداکثر درست نمایی برای پیدا کردن ترکیبی از مقادیر پارامترها که تابع درست نمایی را حداکثر کند، نیاز به انجام حداکثرسازی چند بعدی مقید محدود شده است که در اغلب موارد این موضوع یک مسئله آنالیز عددی مشکل بوده، نیازمند انجام محاسبات زیاد است، (ج) علاوه بر این در روش حداکثر درست نمایی، عموماً محاسبه فواصل اطمینان و آزمون‌های معنی‌داری شامل تقریب‌هایی هستند که این تقریب‌ها اغلب برای نمونه‌های بزرگ دقیق هستند، به عنوان مثال داشتن توزیع احتمال نرمال برای برآوردگر حداکثر درست نمایی در مورد نمونه‌های بزرگ دقیق است، اما در استنباط بیزی، که در آن توزیع پیشین به صورت خودکار اعمال کننده قیدهای پارامتری است، استنباط در مورد مقادیر پارامترها به جای حداکثر سازی براساس توزیع پسین انجام می‌شود که معمولاً نیازمند انتگرال گیری (به عنوان مثال، برای محاسبه میانگین) بوده و تقریب سازی دخالت زیادی در آن ندارد، (د) علاوه بر این روش‌های عددی که در طی دهه ۱۹۵۰ و با استفاده از روش‌های زنجیره مارکف مونت کارلو (MCMC) توسعه و در رایانه‌های توانمند جدید به کار برده شده‌اند، ارزیابی احتمال‌های پسین بیزی را به مقدار زیادی امکان‌پذیر نموده، باعث شده‌اند تا اینکه مشکلات محاسباتی مورد نیاز مدل‌های ژنتیکی پیچیده، که مانعی در برابر تجزیه و تحلیل داده‌های ژنتیکی با استفاده از روش حداکثر درست نمایی یا دیگر روش‌های کلاسیک بودند، مرتفع شود. این موضوعها مهمترین دلایل ایجاد جریان سریع اخیر در مشهور شدن استنباط بیزی در بیشتر شاخه‌های علمی است. در اینجا مثال‌هایی را برای کاربرد استنباط بیزی ارائه می‌دهیم که در آنها این استنباط علاوه بر این که این امکان را مهیا می‌کند که اطلاعات پیشین به صورت کاراتری دخالت داده شوند همزمان مطالعه مدل‌های پیچیده و برآورد پارامترهای زیستی مورد نظر را نیز تسهیل می‌کنند.

ژنتیک جمعیت

علم ژنتیک جمعیت دارای میراث نظری سرشاری است که خاستگاه آن کارهای فیشر^۱، هالدن^۲ و رایت^۳ است. روش-

¹ Fisher

² Haldane

³ Wright

⁴ Coalescent theory

⁵ Parametric Bootstrapping

ژنتیکی» تا ردیابی منابع جمعیتی شیوع پراکندگی یا همه-گیری‌های برطرف شده را شامل می‌شود.

اخیراً، روش‌های انتساب فردی در چندین جهت جدید توسعه پیدا کرده است. بیشتر این کاربردهای جدید، عمدتاً بر روش‌شناسی بیزی و فناوری‌های MCMC متکی است. به طور ویژه، چندین روش بیزی جدید بدین منظور ایجاد شده‌اند که امکان استنباط ترکیبی، شامل تفکیک افراد به زیر جمعیت‌ها و انتساب انفرادی اجداد مهاجر آن‌ها، امکان پذیر شود. در روش دیگری که اخیراً ایجاد شده است، هدف استنباط توأم وجود زیر جمعیت‌ها در داخل جمعیت بزرگ و برآورد شاخص‌های تثبیت مرسوم (آماره F) در بین و در داخل زیر جمعیت‌های مشخص شده، می‌باشد. در نهایت، برای استنباط در مورد نرخ‌های مهاجرت کوتاه مدت (در طی چندین نسل گذشته) با استفاده از ژنوتیپ-های چند جایگاهی افراد، یک روش MCMC بیزی ایجاد شده است. این روش، با دخالت دادن ضریب همخوانی جداگانه هر یک از جمعیت‌ها (ضریب همخوانی برآورد شده به صورت بخشی از روش استنباط MCMC) امکان بررسی انحراف از تعادل هاردی-واینبرگ (یعنی نسبت‌های ژنوتیپی مورد انتظار تحت آمیزش تصادفی) را نیز در داخل جمعیت‌ها امکان پذیر می‌نماید. پیچیدگی‌های چند بعدی این مدل‌ها باعث می‌شود تا انجام استنباط حداکثر درست‌نمایی مشکل شده و از این‌رو روش‌های قابل مقایسه حداکثر درست‌نمایی توسعه پیدا نکرده است. آزمون‌های انتساب چند جایگاهی در حال حاضر در دوران طفولیت خود قرار دارند، اما انتظار می‌رود که در سال‌های آتی به عنوان یک ابزار معمولی مورد استفاده زیست‌شناسان در زمینه‌های مختلفی همانند همه‌گیرشناسی، نقشه‌یابی ژن‌ها در انسان و اکولوژی رفتاری قرار گیرد.

ردیابی انتخاب. هر دو روش مقایسه‌ای و ژنتیک جمعیتی می‌توانند در تعیین جایگاه‌های ژنی کاندیدای تحت تأثیر انتخاب، مورد استفاده قرار گیرد. در تجزیه و تحلیل‌های ژنتیک جمعیتی، استفاده از مدل‌های جمعیت‌شناسی بیزی سلسله‌مراتبی ایده مطرح می‌باشد چرا که در آن این امکان وجود دارد که پارامترهای جمعیت‌شناسی، برای تقلید اثرات انتخاب، در جایگاه‌های ژنی متفاوت در نظر گرفته شوند. چنانچه، احتمال پسین واریانس صفر برای پارامترهای جمعیت‌شناسی جایگاه‌های ژنی، خود نیز

انجام شود. با استفاده از توزیع‌های پسین حاشیه‌ای، تجزیه و تحلیل‌های بیزی به راحتی و با انعطاف به این مسئله می‌پردازد. در حالی‌که در روش‌های کلاسیک جایگزین، یا از برآوردهای نقطه‌ای برای دیگر پارامترهای استفاده می‌کنند و یا اینکه با استفاده از درست‌نمایی‌های نیم‌رخ، فواصل اطمینان برای آن‌ها ایجاد می‌شود. با این‌حال، در استنباط-های جمعیت‌شناسی، توابع درست‌نمایی خیلی پیچیده بوده و تقریب‌های مورد نیاز ایجاد فواصل اطمینان فراوانی-گراها، احتمالاً دقیق نبوده، به کاربردن آن‌ها با توجه به تعداد زیاد پارامترها، از نظر فنی سخت است. تفاوت مابین جایگاه‌های ژنی از نظر پارامترهایی همانند نرخ جهش، می‌تواند با استفاده از مدل‌های بیزی سلسله‌مراتبی^۱، که برای آن‌ها روش‌های جایگزین کلاسیک به آسانی وجود ندارد، مورد بررسی قرار گیرد. در نتیجه توانمندی‌های ایجاد شده در سال‌های اخیر، کاربرد تجزیه و تحلیل بیزی در استنباط‌های جمعیت‌شناسی خیلی شایع شده است. مشکلات محاسباتی می‌توانند با توسعه کارایی روش‌های MCMC و نیز از طریق استفاده از جایگزین‌های MCMC، مرتفع گردند. یک مثال معروف از روش‌های گروه دوم روش محاسباتی بیز تقریبی^۲ (ABC) می‌باشد که در مقام مقایسه با روش‌های MCMC، در مسائل یکسان، می‌تواند با مقدار خیلی کمی اشتباه تا ۱۰۰۰ برابر سریعتر بوده و عملکرد خوبی داشته باشند.

روش‌های انتساب بیزی. مطالعه تفاوت جمعیت‌ها با استفاده از نشانگرهای ژنتیکی دارای تاریخ طولانی است (در کاوالی-اسفورزا^۳ و همکاران مرور شده است). با این-حال، تقریباً اخیراً روش‌هایی برای انتساب افراد به جمعیت‌ها بر اساس ژنوتیپ‌های چند جایگاهی (روش‌های انتساب) توسعه پیدا کرده است. معادله اساسی که روش-های انتساب مورد استفاده قرار دهند، احتمال ژنوتیپ چند جایگاهی یک فرد را، به شرط فراوانی‌های آلی جایگاه‌های ژنی مختلف در جمعیت‌های مختلف، محاسبه می‌کند (در جدول ۱). ثابت شده است که دامنه کاربردهای عملی چنین آزمون‌های انتساب، گسترده می‌باشد. این کاربردها، از ردیابی اختلاط جمعیتی پنهان در «مطالعات پیوستگی

¹ Hierarchical Bayesian Models

² Approximate Bayesian Computation

³ Cavalli-Sforza

زنجیره‌های مارکف یا مدل‌های مارکف پنهان^۲ (HMMs) هسته اصلی بیشتر روش‌های تحلیل توالی مبتنی بر حداکثر درست‌نمایی می‌باشد. این روش‌ها برای پیدا کردن جواب‌های حداکثر درست‌نمایی با ابعاد بالا از برنامه نویسی پویا استفاده می‌کنند. برخی تحلیل‌های مبتنی بر درست‌نمایی، تابع امتیازبندی را تولید می‌کنند که شامل محاسبات بیزی است. برای مثال، نرم افزار GeneMark که برای حاشیه نویسی ژنوم پروکاریوت‌ها استفاده می‌شود، میزان درست‌نمایی را تحت شرایط مختلف محاسبه (احتمال داده‌ها را به شرط اینکه آن بخش مربوط به بخش کد شوند، غیر کد شوند و الی آخر باشد) و سپس یکسری محاسبات بیز تجربی برای گزینش بین آن‌ها، همانند چیزی که در قبل برای ردیابی نوکلئوتیدهای تحت انتخاب توضیح داده شد، انجام می‌شود.

گروه بزرگی از تحلیل‌های بیزی از مدل‌هایی منشاء گرفته‌اند که در آن‌ها فرض می‌شود بازهای جایگاه‌های نوکلئوتیدی و اسیدهای آمینه‌ای قرار گرفته، متغیرهای تصادفی مستخرج شده از توزیع‌های احتمالی هستند که این توزیع‌ها برای مناطق مختلف متفاوت است. در این شرایط مسئله استنباط در کنار پیدا کردن پارامترهایی همانند ترکیب بازی داخل و بیرون مناطق، یافتن مناطق مختلف نیز خواهد بود. از این نظر، روش‌های بیزی در ابتدا برای مدل‌سازی ترازبازی پروتئینی^۳ به کار رفت، روشی که به تراز بندی محلی توسعه یافته و برای تعیین مکان‌های اتصال، عامل ترانسکریپتوم نیز مورد استفاده قرار گرفته است. مدل‌بندی بیزی براساس این روش برای به دست آوردن توزیع‌های حاشیه‌ای نقاط تغییر (کران‌های مناطق) و ترکیب بازی در طول یک توالی نیز به کار رفته است. روش‌های حداکثر درست‌نمایی در مسائلی از این قبیل، عموماً از نظر تعداد پارامترهایی که می‌توانند در نظر بگیرند، محدودند و در اغلب موارد به خاطر بهینه سازی با ابعاد بالای مورد نیاز، آزمون معنی داری نیز در این روش، محدود است. در مقابل، روش بیزی امکان در نظر گرفتن پارامترهای بیشتر را فراهم (روش بیزی ضرورتاً این امکان را بوجود می‌آورد که پارامترهایی که در روش‌های حداکثر درست‌نمایی ثابت در نظر گرفته می‌شوند در این روش

نزدیک صفر باشد، این احتمال وجود دارد که برخی از این جایگاه‌های ژنی تحت انتخاب بوده باشند. برای تعیین کاندیدها در انتخاب تطبیفی در جمعیت‌های تقسیم شده نیز روش مشابه مورد استفاده قرار گرفته است. روشی نیز برای پیدا کردن توزیع اثرات انتخابی در بین جایگاه‌های ژنی توضیح داده شده است.

روش‌های ژنتیک جمعیتی ردیابی انتخاب می‌بایست نسبت به مدلی که برازش می‌شود، حساس باشند، چرا که وقایع جمعیت شناختی همانند، تنگناگاه^۱ جمعیتی، ممکن است تشدید کننده یا پوشاننده اثرات انتخاب باشند. با استفاده از داده‌های توالی‌یابی به دست آمده از گونه‌های مختلف، که در آن‌ها اثرات جمعیت شناختی به خاطر اینکه تفرق آلل‌ها در داخل جمعیت در نظر گرفته نمی‌شوند مهم نیست، استنباط‌های توانمندتری در دسترس اند. تجزیه و تحلیل‌های این سطح، تمرکزشان بر نسبت w ، یعنی نسبت جایگزینی‌های نوکلئوتیدی غیر تغییر دهنده اسیدآمینه در پروتئین به جایگزین‌های منجر به تغییر اسیدهای آمینه در پروتئین، است. چنانچه تمامی جایگزینی‌های اسیدآمینه‌ای خنثی باشند، این نسبت باید برابر یک و چنانچه آن‌ها زیان‌آور باشند، این نسبت باید کم‌تر از یک و اگر سودمند باشند (انتخاب مثبت) این نسبت باید بیشتر از یک باشد. براساس این اصول، روش بیزی به منظور تعیین اینکه کدام کدهای داخل ژن تحت انتخاب مثبت بوده‌اند، مورد استفاده قرار گرفته است. در این روش (یک روش بیز تجربی) برآوردهای نقطه‌ای روش حداکثر درست‌نمایی برای پارامترهای فیلوژنتیک، برای محاسبه احتمال پسین اینکه یک کد متعلق به کدام یک از این سه گروه ($w = 0/1$) یا $w >$ می‌باشد) است مورد استفاده قرار می‌گیرد. روش فیلوژنتیک بیزی برآوردهای خیلی کامل‌تری را برای این احتمال ارائه می‌دهد.

ژنومیکس

تحلیل توالی. جنبه‌های غیر فیلوژنتیکی تحلیل توالی دارای تاریخ گسترده و غنی از روش‌های مبتنی بر مدل بوده، شامل اولین کاربرد روش MCMC برای یک مسئله زیست‌شناسی بوده است.

² Hidden Markov Model

³ Protein Alignment

¹ Bottleneck

پیشین مختلف بر تعداد زیادی از پارامترهای مدل به این آسانی امکان پذیر نیست. دیگر موضوع مطرح در مطالعات مبتنی بر MCMC، مشکل بودن ارزیابی همگرایی آن بویژه برای مدل‌هایی با تعداد متغیر ابعاد می‌باشد. عموماً، بیشتر روش‌های بیزی آهسته هستند که این یک مانع بزرگ برای هر چیزی بیش از بررسی مدل اولیه می‌باشد.

روندهای جاری حاکی از این هستند که روش‌های استاندارد تصحیح شده MCMC به صورت فزاینده‌ای در حال کشف شدن هستند. در مطالعاتی که در آن‌ها تعداد زیادی از پارامترها مورد نظر محقق نمی‌شوند (همانند تاریخ شجره‌ای در مدل‌های ژنتیک جمعیتی) و تنها تعداد اندکی از آن‌ها مورد توجه هستند، روش ABC بطور ویژه-ای امید بخش به نظر می‌رسد. از آنجایی که این روش جذاب و به کار گیری آزادانه است، به عنوان مثال برای زیست شناسانی که از شبیه‌سازی کامپیوتری لذت می‌برند اما اطلاعات پایه‌ای اندکی در علم احتمالات دارند، باعث می‌شود تا شبیه‌سازی به یک ابزار مطلوب برای استنباط تبدیل شود.

دیگر حوزه در حال رشد که در این مرور به آن پرداخته نشد، عبارت از استفاده از شبکه‌های بیزی برای ترکیب کردن نتایج حاصل از تحلیل‌های مختلف در مجموعه داده‌های یکسان است. با این حال، چنین استدلال می‌شود که چنین روش‌هایی اگر چه مفید بوده، از نظر تجاری دارای امتیازاتی هستند اما فناوری‌هایی هستند که به آسانی در تحقیقات علمی به کار نمی‌روند. در مقابل روش‌های توضیح داده شده در اینجا مبتنی بر مدل‌های احتمالی فرآیندهایی هستند که به صورت یک الگو در در آمده‌اند. آن‌ها پارامترهایی دارند که مرتبط با کمیت‌ها بوده و می‌توانند به صورت اصولی اندازه‌گیری و مورد آزمون قرار گیرند. در حال حاضر، انقلاب بیزی در مرحله اولیه خودش بوده و تا فرو نشستن گردوغبار انقلاب و قضاوت ما در مورد این‌که راه نجات بخش برای کشف چه راهی است، هنوز ادامه خواهد داشت.

می‌توانند تغییرات داشته باشند) کرده و استنباط کامل در مورد هر پارامتر را به وجود آورد و از طریق انتخاب «مدل آزمون»، معنی‌داری را نیز میسر می‌کند. در اغلب موارد دخالت دادن مدل HMM در چارچوب MCMC، به راحتی ممکن بوده و بنابراین احتمال توسعه تحلیل بیزی مبتنی بر چارچوب حداکثر درست نمایی برای تحلیل داده‌های توالی در آینده زیاد است. از کاربردهای دیگر استنباط بیزی می‌توان به «تشخیص SNP»، «استنباط بیزی هاپلوتیپ با استفاده از نمونه‌های جمعیتی»، «استنباط سطوح بیان ژن»، «علم ژنتیک انسانی» و «نقشه یابی دقیق ژن‌های حساس پایه‌ای» اشاره نمود.

دورنماها و پیش بینی‌ها

انعطاف پذیری زیاد روش‌های بیزی که در مثال‌های این مقاله نشان داده شد، حاکی از نیاز به آزمون‌های مدل توانمندتر نیز هست. در استنباط فراوانی‌گراها، کار معمول عبارت از شبیه سازی تعداد زیادی (هزاران) از گروه‌های داده‌ای آزمون، که در آن‌ها مقادیر پارامتر حقیقی شناخته شده بوده، و سپس اندازه‌گیری اریب، میانگین توان دوم خطا و میزان پوشش برآوردها می‌باشد. چنین روشی به آسانی در داخل مدل بیزی قرار نمی‌گیرد، اما در اغلب موارد روش ساده برای مقایسه با روش‌های فراوانی‌گراها بوده است. برای ارزیابی مدل در استنباط بیزی، پیشنهاد شده است که پارامترها از توزیع پسین استخراج شده و سپس برای شبیه سازی دیگر مجموعه‌های داده‌ای مورد استفاده قرار گیرند. این همان توزیع پیش بینی کننده پسین، یعنی توزیع دیگر مجموعه‌های داده‌ای به شرط داده‌های مشاهده شده، می‌باشد. برای بررسی این‌که آیا مدل مورد نظر یک مدل معقول است یا نه، خلاصه آماره‌های اندازه‌گیری شده در داده‌های واقعی می‌تواند با خلاصه آماره‌های داده‌های شبیه‌سازی مورد مقایسه قرار گیرد. با این حال، این روش در عمل، خیلی نادر بکار گرفته شده است. همین‌طور اگرچه بررسی میزان حساسیت مدل به توزیع‌های پیشین، مهم می‌باشد، اما در مدل‌های سلسله مراتبی پیچیده، معمولاً بررسی اثرات سیستماتیک توزیع‌های

منابع

- Beaumont, M. A., & Rannala, B. (2004). The Bayesian revolution in genetics. *Nature Reviews. Genetics*, 5(4), 251.
- Shoemaker, J. S., Painter, I. S., & Weir, B. S. (1999). Bayesian statistics in genetics: a guide for the uninitiated. *Trends in Genetics*, 15(9), 354-358.