معرفی پیشین فرآیند دیریکله در چارچوب مدل‌های بیزی ناپارامتری

عاطفه جاویدی آل‌سعدی [۱]، سمیه راهپیمایی کازرونی [۲]، مجید جعفری خالدی [۳]

چکیده:

مدل‌های آماری برای شناخت فرآیندی که داده‌ها از آن تولید شده، استفاده می‌شود. در بیشتر مدل‌ها فرض می‌شود متغیرهای تصادفی Y_i، $i = 1, ..., n$، نمونه‌ای تصادفی از توزیع F هستند، که F متعلق به یک کلاس از خانواده توزیع‌های پارامتری است. اما در عمل همیشه نمی‌توان انتظار داشت که یک مدل پارامتری برای توصیف داده‌ها مناسب باشد. در این شرایط می‌توان مدل‌های پارامتری را کنار گذاشت و از مدل‌های انعطاف‌پذیر و نیرومندتری برای تحلیل داده‌ها استفاده کرد. در چارچوب روش بیز ناپارامتری با تعریف یک توزیع پیشین روی فضای کل توزیع‌های احتمالی و در نظر گرفتن آن برای توزیع متغیر تصادفی این انعطاف‌پذیری حاصل می‌شود. بعبارت دیگر فرآیندهای تصادفی روی خانواده‌ای از توابع توزیع تعریف می‌شوند و بعنوان پیشین برای توزیع تصادفی به‌کار می‌روند. از جمله مهم‌ترین این پیشین‌ها فرآیند دیریکله است که به دلیل برخورداری از ویژگی‌های مهم و جالب توجه در گستره وسیعی از مسائل بیز ناپارامتری مورد استفاده قرار می‌گیرد. در این مقاله این فرآیند و خواص آن معرفی می‌شود.

واژه‌های کلیدی: توزیع دیریکله، فرآیند دیریکله، بیز ناپارامتری.

۱ مقدمه

فرض کنید داده‌ها تحققی از متغیرهای تصادفی $Y_1, \cdots, Y_n$ هستند. معمولاً فرض می‌شود که $Y_1, \cdots, Y_n$ به‌طور مستقل از توزیع احتمالی $\mu(B) = P(Y \in B)$ تولید شده است، که در آن B زیر مجموعه‌ای اندازه‌پذیر از فضای نمونه‌ای و $P(A)$ احتمال پیشامد A را به نمایش می‌گذارد. فرض کنید $f(x)$ تابع چگالی متناظر با اندازه احتمال $\mu(\cdot)$ باشد. هنگامی که این توزیع نامعلوم است، در مدل‌های پارامتری[۴] معمولاً فرض می‌شود که $f(x)$ به‌صورت $f_{\boldsymbol{\theta}}(x)$ بوده و در واقع یک عضو از خانواده توزیع‌های $F = \{f_{\boldsymbol{\theta}} : \theta \in \Theta\}$ است. مدل‌های پارامتری روی فضای متناهی بعد[۵] $\Theta \subseteq \mathbb{R}^p$ تعریف می‌شود، که در آن p یک عدد مثبت است. لازم به ذکر است که در این حالت فضای پارامتری یک زیرمجموعه از فضای خطی متناهی بعد S است.

تعریف ۱.۱. فضای خطی S متناهی بعد است اگر دارای تعداد متناهی عنصر در فضا باشد. در واقع S یک فضای متناهی بعد است اگر عناصر $\theta_1, ..., \theta_m \in S$، که m یک عدد مثبت متناهی است، موجود باشد به‌گونه‌ای که هر عنصر $\theta \in S$ یک ترکیب خطی از $\theta_1, ..., \theta_m$ باشد. بعد یک فضای خطی متناهی بعد به‌صورت تعداد عناصر مستقل خطی که در فضای کل گسترده شده است، تعریف می‌شود.

یک هدف مهم در مدل‌های پارامتری، برآورد یک مقدار قابل قبول برای پارامتر مدل، یعنی θ، است. همان‌طور که اشاره شد در مدل‌های پارامتری فرض محدودکننده معلوم بودن فرم توزیع ممکن است نتایج استنباط آماری را نامعتبر سازد. اما می‌توان با کنار گذاشتن مفروضات پارامتری مدلی انعطاف‌پذیر و نیرومندتر ارائه کرد. این‌گونه مدل‌ها با در نظر گرفتن حالتی که در آن تابع چگالی در فضای بزرگتر و پارامترها دارای بعد نامتناهی هستند،

[۱] گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس

[۲] گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس

[۳] عضو هیئت علمی گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس

[۴] Parametric Models

[۵] Finite Dimensional Space

به‌وجود می‌آیند.

تعریف ۲.۱. فرض کنید $S \subseteq \Theta$ باشد. فضای خطی S که شامل تعداد نامتناهی عنصر است فضای پارامتری نامتناهی بعد[۶] و پارامترهای متعلق به این فضا، پارامترهای نامتناهی بعد نامیده می‌شود.

فضای توابع پیوسته، $\{m(\cdot)\}$ یک تابع پیوسته $S = \{m(x); x \in \mathbb{R}\}$ مثالی از فضای خطی نامتناهی بعد است. مدل‌هایی که تنها پارامترهای نامتناهی بعد را در نظر می‌گیرند به عنوان مدل‌های ناپارامتری شناخته می‌شوند (گوش و رامامورتی، ۲۰۰۳؛ تسیاتیس، ۲۰۰۶). در برخی حالت‌ها، پارامترهای مدل را به دو دسته تقسیم می‌کنند، یعنی θ به‌صورت (θ_1, θ_2) بازنویسی می‌شود. در این پاراموتربندی θ_1 پارامتری با بعد q و θ_2 پارامتری نامتناهی بعد فرض می‌شود. این مدل را به‌دلیل وجود یک جزء پارامتری و یک جزء ناپارامتری مدل نیمه‌پارامتری می‌نامند (تسیاتیس، ۲۰۰۶). در واقع مدل نیمه‌پارامتری ترکیبی از مدل‌های پارامتری و ناپارامتری است. به‌طور کلی در مدل‌های بیز ناپارامتری برای تحلیل داده‌های Y_i، $i = 1, ..., n$، که به‌طور مستقل از توزیع نامعلوم G تولید شده‌اند، فرض می‌شود که G توزیعی تصادفی بوده و برای این توزیع تصادفی پیشینی به‌صورت $\pi(G)$ در نظر می‌گیرند. در واقع داریم:

$$Y_1, \ldots, Y_n \mid G \overset{iid}{\sim} G$$
$$G \sim \pi(G).$$

در سالهای اخیر مدل‌های آماری مختلف بر اساس روش بیز ناپارامتری مورد بررسی قرار گرفته‌اند. در ادامه روش بیز ناپارامتری در یک مدل رگرسیونی بیان می‌شود. یک مدل رگرسیونی با متغیرهای وابسته y_i و متغیر کمکی x_i را به‌صورت $y_i = f(x_i) + \epsilon_i$، که $\epsilon \sim p_\epsilon(\epsilon_i)$ در نظر بگیرید. در مدل‌های رگرسیون پارامتری توزیع خطاها و $f(x_i)$، پارامتری در نظر گرفته می‌شوند. اما در مدل‌های رگرسیون بیزی ناپارامتری آنها را ناپارامتری در نظر می‌گیرند که منتهی به سه حالت زیر می‌شود:

۱. مدل خطای ناپارامتری: در این مدل فرض می‌شود که ϵ_i دارای توزیع تصادفی G است، یعنی $\epsilon_i \sim G$، که G دارای پیشین $\pi(G)$ است و $f(x_i)$ یک تابع پارامتری در نظر گرفته می‌شود.

۲. مدل تابع میانگین ناپارامتری: در این مدل تابع میانگین دارای توزیعی تصادفی در نظر گرفته می‌شود و پیشینی برای این توزیع انتخاب می‌شود. به‌عبارت دیگر $f \sim \pi(f)$ است که $\pi(f)$ یک توزیع پیشین است. پیشین مرسوم برای f فرآیندهای گاوسی است (مولر، ۲۰۱۳).

۳. رگرسیون ناپارامتری کامل: در این مدل هم تابع میانگین و هم توزیع خطاها تصادفی فرض می‌شود و برای آنها توزیع‌های پیشینی در نظر گرفته می‌شود.

برای دیدن جزئیات و مثال‌های بیشتر پیرامون روشهای بیز ناپارامتری می‌توان به مولر و میترا (۲۰۱۳) مراجعه کرد.

شیوه‌های مختلفی برای تعریف پیشین برای توزیع تصادفی وجود دارد (دی و همکاران، ۱۹۹۸). رایج‌ترین و پرکاربردترین آن‌ها بر مبنای فرآیندهای تصادفی است. فرآیندهای تصادفی روی خانواده‌ای از توابع توزیع تعریف می‌شوند و می‌تواند به عنوان یک پیشین مناسب برای توزیع تصادفی باشد. به دلیل پیچیدگی محاسبات در روش بیز ناپارامتری انتخاب نوع فرآیند تصادفی خود مسئله‌ی مهمی است. فرآیندی باید انتخاب شود که انعطاف‌پذیر بوده و دارای این ویژگی باشد که محاسبات در چارچوب آن راحت‌تر و سریع‌تر انجام شود، به عبارت دیگر یک پیشین مزدوج باشد. از جمله این فرآیندها، فرآیند دیریکله است. فرگوسن (۱۹۷۳) فرآیند دیریکله را به عنوان یک توزیع پیشین روی فضای تمام توابع توزیع تعریف کرد، یک پیشین ناآگاهی‌بخش که موجب راحتی محاسبات و همچنین استنباط دقیق و کارا می‌شود. توسعه و گسترش روش‌های محاسبات بیزی، مانند روش‌های $MCMC$ منجر به استفاده از فرآیند دیریکله به عنوان توزیع پیشین در گستره وسیعی از مسائل بیز ناپارامتری، رگرسیون ناپارامتری، برآورد چگالی ناپارامتری و مدل‌های با اثرات تصادفی گردید. در این مقاله فرآیند دیریکله به عنوان یک پیشین شناخته شده معرفی می‌شود. برای این منظور ابتدا در بخش ۲ توزیع دیریکله و ویژگی‌های آن را معرفی می‌کنیم. در بخش ۳ به معرفی فرآیند دیریکله و ویژگی‌های

[۶] Infinite Dimensional Space

آن می‌پردازیم. در ادامه در بخش ٤ ضمن معرفی مدل‌های مبتنی بر فرآیند دیریکله با ویژگی‌ها و کاربردهای هر یک از آن‌ها تا حدودی آشنا می‌شویم و در نهایت در بخش ۵ فرآیند دیریکله را در ساختار مدل‌های متغیر پنهان به کار می‌بریم.

۲ توزیع دیریکله

توزیع دیریکله در نظریه احتمال و آمار یک توزیع پیوسته است. در واقع این توزیع یک توزیع چند پارامتری تعمیم‌یافته از توزیع بتا می‌باشد. معمولاً از توزیع دیریکله به عنوان توزیع پیشین در استنباط بیزی استفاده می‌شود، چرا که این توزیع یک پیشین مزدوج[۷] برای پارامترهای توزیع چند جمله‌ای و توزیع رسته‌ای است.

تعریف ۱.۲. فرض کنید $\Theta = \{\theta_1, \cdots, \theta_n\}$ یک توزیع احتمال روی فضای گسسته $\chi = \{\chi_1, \cdots, \chi_n\}$ و X متغیر تصادفی روی χ باشد به‌طوری‌که $p(X = \chi_i) = \theta_i$. توزیع دیریکله روی Θ با پارامترهای $\boldsymbol{\beta} = \{\beta_1, \cdots, \beta_n\}$ به صورت

$$p(\Theta \mid \boldsymbol{\beta}) = \frac{\Gamma(\beta_0)}{\prod_{i=1}^{n} \Gamma(\beta_i)} \prod_{i=1}^{n} \theta_i^{\beta_i - 1}$$

می‌باشد، که در آن $\theta_i > 0 \ , \sum_{i=1}^{n} \theta i = 1$ و همچنین $\beta_0 = \sum_{i=1}^{n} \beta_i \ , \beta_i > 0$.

گشتاورهای این توزیع عبارت از:

$$E(\theta_i) = \frac{\beta_i}{\beta_0},$$
$$Var(\theta_i) = \frac{\beta_i(\beta_0 - \beta_i)}{\beta_i^2(\beta_0 + 1)},$$
$$Cov(\theta_i, \theta j) = \frac{-\beta_i \beta j}{\beta_i^2(\beta_0 + 1)},$$

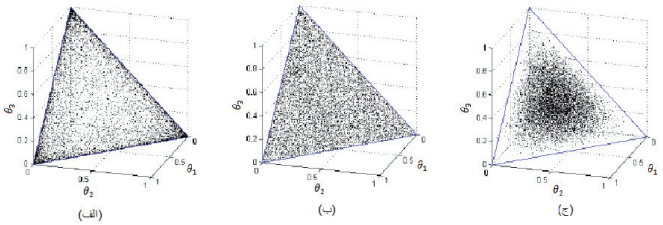
هستند. باید به این نکته توجه شود که وجود کوواریانس نشان‌دهنده وجود وابستگی بین داده‌های تولید شده از این توزیع است. یک پارامتربندی جایگزین برای بردار پارامترهای β به صورت αm_i می‌باشد که در آن $\alpha = \beta_0$ پارامتر دقت یا تمرکز[۸] است و پراکندگی توزیع را حول میانگین کنترل می‌کند و $m_i = \frac{\beta_i}{\beta_0}$ یک اندازه پایه و معیار میانگین است. در این حالت:

$$p(\Theta \mid \alpha m_1, \ldots, \alpha m_n) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{n} \Gamma(\alpha m_i)} \prod_{i=1}^{n} \theta_i^{\alpha m_i - 1} \quad (۱)$$

رابطه (؟؟) یک توزیع دیریکله را به نمایش می‌گذارد و می‌گوییم $\Theta \sim Dir(\alpha M)$، که در آن $M = \{m_1, ..., m_n\}$ با ویژگی $\sum_{i=1}^{n} m_i = 1$ است. در این حالت خواهیم داشت:

$$E(\theta_i) = m_i$$
$$Var(\theta_i) = \frac{m_i(1 - m_i)}{\alpha(\alpha + 1)}$$

با توجه به رابطه فوق، با افزایش α، واریانس کاهش یافته و داده‌ها اطراف میانگین متمرکز می‌شوند. شکل ۱ بر اساس شبیه سازی با $n = 10000$ از یک توزیع دیریکله سه بعدی با M یکنواخت و $\alpha = 1, 3, 10$ به تصویر کشیده شده است. همان‌طور که از این شکل مشاهده می‌شود هنگامی که $\alpha = 3$، یعنی برابر با بعد توزیع دیریکله است، داده‌ها به‌صورت کاملاً یکنواخت پراکنده می‌شوند. هنگامی که $\alpha < 3$ است، داده‌ها از میانگین دور و در حاشیه تمرکز دارند. در مقابل هنگامی که $\alpha > 3$، داده‌ها حول میانگین، یعنی M متمرکز می‌شوند. توزیع دیریکله همانند توزیع بتا علاوه بر حالت‌های متقارن، چولگی‌ها را در بر می‌گیرد، که این خاصیت نشان دهنده انعطاف‌پذیری این توزیع است.



شکل ۱. توزیع دیریکله ۳ بعدی با M یکنواخت و (الف)$\alpha = 1$، (ب)$\alpha = 3$ و (ج)$\alpha = 10$ بر اساس ۱۰۰۰۰ نمونه تولید شده.

برخی ویژگی‌های این توزیع عبارتند از:

۱. برای $n = 2$ توزیع دیریکله، یک توزیع بتا است. بعلاوه توزیع حاشیه‌ای دیریکله یک توزیع بتا با پارامترهای $(\alpha m_i, \alpha - \alpha m_i)$ می‌باشد (فارو ملکلم، ۲۰۰۸).

۲. توزیع دیریکله یک پیشین مزدوج برای پارامترهای توزیع چند جمله‌ای و رسته‌ای است. به بیان دیگر: فرض کنید

[۷] Conjugate Prior

[۸] Concentration Parameter

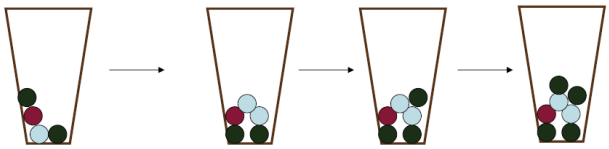
N مشاهده $x_1, ..., x_N$ از یک توزیع گسسته با پارامتر $\Theta = \{\theta_1, ..., \theta_n\}$ داریم. اگر n_i، $i = 1, ..., n$ ، تعداد دفعاتی باشد که χ_i در N مشاهده رخ می‌دهد و $\sum_{i=1}^{n} n_i = N$ باشد، با در نظر گرفتن توزیع پیشین دیریکله برای Θ، توزیع پسین Θ با استفاده از قانون بیز به صورت

$$
\begin{aligned}
& p(\Theta \mid \alpha, M, X_{1:N}) \\
& \propto \prod_{n=1}^{N} p(X_n \mid \alpha, M, \Theta) p(\Theta \mid \alpha, M) \\
& \propto \prod_{i=1}^{n} \theta_i^{n_i} \prod_{i=1}^{n} \theta_i^{\alpha m_i - 1} \\
& \propto \prod_{i=1}^{n} \theta_i^{\alpha m_i + n_i - 1} \\
& \propto Dir(\alpha m_1 + n_1, \ldots, \alpha m_n + n_n)
\end{aligned}
$$

است. همان طور که ملاحظه می‌شود با انتخاب پیشین دیریکله توزیع پسین نیز یک توزیع دیریکله است.

۱.۲ نمونه‌گیری از توزیع دیریکله

یک روش تولید نمونه از توزیع دیریکله استفاده از ظرف پولیا[۹] است. فرض کنید می‌خواهیم از توزیع دیریکله $\Theta \sim Dir(\alpha M)$ نمونه تولید کنیم. برای شروع فرض کنید α گوی با رنگ‌های مختلف در ظرف وجود دارد، به‌طوری‌که αm_1 گوی از رنگ اول، αm_2 گوی از رنگ دوم و الی آخر. به طور تصادفی یک گوی از ظرف خارج می‌کنیم، سپس گوی را همراه با یک گوی هم‌رنگ آن به ظرف برمی‌گردانیم. بلکول و مک‌کویین (۱۹۷۳) نشان دادند که با تکرار این روش در دفعات زیاد، نسبت رنگ هر گوی در کیسه همگرا به توزیع دیریکله است.



شکل ۲. مراحل تولید داده از یک توزیع دیریکله بر اساس ظرف پولیا

اگر احتمال اینکه در گام i–ام گوی به رنگ j–ام خارج شود را به صورت $p(\theta_i = j)$ نشان دهیم داریم:

$$
\begin{aligned}
& p(\theta_1 = j) = \frac{\alpha m_j}{\sum_{i=1}^{n} \alpha m_i}, \\
& p(\theta_2 = j \mid \theta_1) = \frac{\alpha m_j + \delta(\theta_1 = j)}{\sum_{i=1}^{n} \alpha m_i}, \\
& p(\theta_{N+1} = j \mid \theta_{1:N}) = \frac{\alpha m_j + \sum_{i=1}^{n} \delta(\theta_i = j)}{\alpha + N},
\end{aligned}
$$

که در آن N تعداد مشاهدات و δ تابع دلتای دیراک[۱۰] است. شکل ۲ سه مرحله از فرآیند تولید داده از ظرف پولیا با $\alpha = 4$ را به نمایش می‌گذارد.

۳ فرآیند دیریکله

فرآیند دیریکله یک توسیع از توزیع دیریکله در فضای پیوسته می‌باشد. فرض کنید χ یک فضای نمونه، β سیگما میدان بورل مجموعه‌های χ و F فضای تمامی اندازه‌های احتمال تعریف شده روی (χ, β) باشد.

تعریف ۱.۳. فرض کنید α یک مقدار حقیقی مثبت، G_0 یک اندازه متناهی، نامنفی و غیرصفر روی فضای اندازه‌پذیر (χ, β) و $G(\cdot)$ یک فرآیند اندیس‌گذاری شده بوسیله عناصر β است. گوییم G یک فرآیند دیریکله با پارامتر (α, G_0) است و می‌نویسیم $G \sim DP(\alpha G_0)$ هرگاه

۱. $G(B)$ برای $B \in \beta$، یک متغیر تصادفی باشد به‌گونه‌ای که مقادیرش را در بازه $[0, 1]$ می‌گیرد.

۲. هر تحقق از G یک اندازه احتمال روی فضای (χ, β) باشد.

۳. به ازای هر تعداد افرازهای متناهی اندازه‌پذیر $B_1, ..., B_k$، برای $k = 1, 2, ...$، از χ (B_iها اندازه‌پذیر و مجزا بوده و $\bigcup_{i=1}^{k} B_i = \chi$) بردار تصادفی $(G(B_1), ..., G(B_k))$ دارای توزیع دیریکله با پارامترهای $(\alpha G_0(B_1), ..., \alpha G_0(B_k))$ باشد.

در این تعریف α پارامتر دقت[۱۱] و G_0 اندازه مرکزی[۱۲] یا توزیع پایه[۱۳] فرآیند دیریکله نامیده می‌شود.

[۹] Polya Urn

[۱۰] Dirac Delta Function

[۱۱] Precision Parameter

[۱۲] Center Measure

[۱۳] Base Distrubution

وجود فرآیند دیریکله به عنوان یک فرآیند تصادفی تعریف شده روی فضای (χ, β) از طریق قضیه سازگاری کلموگروف (کلموگروف، ۱۹۳۳) توسط فرگوسن اثبات گردید. بلکول (۱۹۷۳) اثبات دیگری را ارائه داده است.

از آن‌جایی که فرآیند دیریکله توزیعی روی اندازه احتمال تصادفی G است، برای هر $B \in \beta$، $G(B)$ یک متغیر تصادفی است. با توجه به تعریف این فرآیند، می‌توان گفت:

$$G(B) \sim Beta\{\alpha G_0(B), \alpha(1 - G_0(B))\}.$$

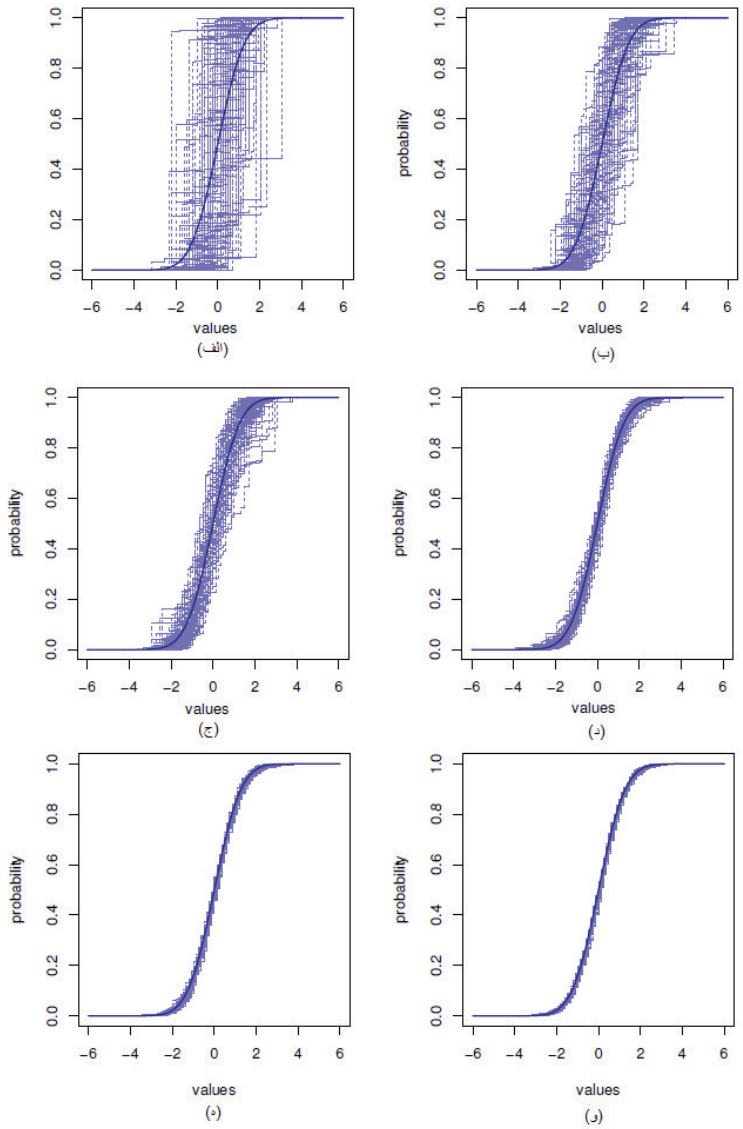
بنابراین

$$E(G(B)) = G_0(B)$$
$$Var(G(B)) = \frac{G_0(B)(1 - G_0(B))}{1 + \alpha}$$

همان‌طور که دیده می‌شود پارامتر دقت، یعنی α، در فرآیند دیریکله انحراف G از G_0 را کنترل می‌کند. با افزایش α، G بیشتر اطراف G_0 متمرکز می‌شود، به‌عبارت دیگر هنگامی که $\alpha \to \infty$، G به سمت G_0 می‌رود. شکل ۳ تاثیر افزایش پارامتر دقت بر G را به نمایش می‌گذارد. در این شکل توزیع پایه تابع توزیع نرمال استاندارد در نظر گرفته شده و ۱۰۰ نمونه از فرایند دیریکله به ازای

$$\alpha = 1, 5, 20, 100, 500, 1000$$

تولید شده است. همان‌طور که مشاهده می‌شود با افزایش پارامتر دقت نمونه‌ها اطراف توزیع مرکزی متمرکز شده‌اند.



شکل ۳. نمودار ۱۰۰ داده تولید شده از $DP(\alpha G_0)$ با توزیع مرکزی نرمال استاندارد و به ترتیب $\alpha = 1, 5, 20, 100, 100, 500, 1000$.
در تمامی نمودارها توزیع پایه G_0 نیز نمایش داده شده است.

لازم به ذکر است که تکیه‌گاه فرآیند همان تکیه‌گاه توزیع مرکزی، یعنی G_0 است. در ادامه به بررسی برخی خصوصیات شناخته شده فرآیند دیریکله بر اساس سه قضیه می‌پردازیم.
هر تحقق از فرآیند دیریکله یک توزیع گسسته است. این موضوع در قالب قضیه زیر مطرح می‌شود:

قضیه ۲.۳. (فرگوسن، ۱۹۷۳) فرض کنید $G \sim DP(\alpha G_0)$ باشد، آنگاه G تقریباً همه جا $(a.s.)$ یک اندازه احتمال گسسته است.

قضیه فوق به روش‌های مختلفی توسط فرگوسن (۱۹۷۳) و بلکول (۱۹۷۳) اثبات گردیده است. فرگوسن این قضیه را از طریق یک نمایش گاما ثابت نمود. بلکول برای اثبات آن از طرح ظرف پولیا استفاده کرد. همچنین این قضیه باعث ایجاد توجه خاصی به فرآیند دیریکله به عنوان یک توزیع پیشین روی توزیع‌های گسسته گردید. بنابراین با فرض آنکه $X_1, ..., X_n$ نمونه‌هایی از یک توزیع تحقق یافته G از یک فرآیند دیریکله باشند. احتمال وجود X_i–های مشابه در بین نمونه‌ها وجود دارد. فرض کنید $n^* < n$ تعداد خوشه‌های متفاوتی باشد که X_i–ها در آن خوشه‌ها قرار می‌گیرند مجموعه مقادیر متفاوت با $X^* = (X_1^*, ..., X_{n^*}^*)$ نشان داده می‌شود. همچنین فرض کنید n_j، $j = 1, ..., n^*$، تعداد نمونه‌های هر خوشه را نمایش می‌دهد. بلکول و مک‌کوئین (۱۹۷۳) نشان دادند که برای بردار نمونه $X_1, ..., X_n$ از یک توزیع تحقق‌یافته فرآیند دیریکله، توزیع پیش‌گو به‌صورت

$$p(X_{n+1} \mid X_1, ..., X_n) \propto \sum_{j=1}^{n^*} n_j \delta_{X_j^*} + \alpha G_0 \qquad (۲)$$

است. این رابطه نشان می‌دهد که نمونه جدید با احتمال متناسب با n_j یک نمونه مشابه نمونه‌های قبلی است یا با احتمال متناسب با α یک نمونه جدید از G_0 است. این رابطه با انتگرال‌گیری نسبت به توزیع تصادفی G حاصل می‌شود. در این حالت نمونه‌ها وابسته و دارای توزیع حاشیه‌ای G_0 هستند. مدل (؟؟) را اصطلاحاً نمایش فرآیند دیریکله از طریق فرآیند ظرف پولیا گویند. یکی از مهم‌ترین خصوصیات فرآیند دیریکله این است که یک پیشین مزدوج است.

قضیه ۳.۳. (فرگوسن،۱۹۷۳) فرض کنید $X_1, \ldots, X_n \mid G \overset{iid}{\sim} G$ و $G \mid \alpha, G_0 \sim DP(\alpha G_0)$ آنگاه توزیع پسین G به شرط $X_1, \ldots, X_n$ به صورت

$$[G \mid X_1, \ldots, X_n, \alpha, G_0] \sim DP(\alpha G_0 + \textstyle\sum_{i=1}^{n} \delta_{X_i})$$

است.

یک نتیجه این قضیه که در محاسبات بیزی کاربرد دارد به صورت زیر است:
فرض کنید $X_1, \ldots, X_n, X_{n+1} \mid G \overset{iid}{\sim} G$ و $G \mid \alpha, G_0 \sim DP(\alpha G_0)$، آنگاه

$$X_{n+1} \mid \alpha, G_0, X_1 \ldots, X_n \sim \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_{X_i} \qquad (۳)$$

ستورامن (۱۹۹٤) تصویر دیگری از فرآیند دیریکله را در قالب قضیه زیر ارائه کرد و بر اساس آن ثابت کرد که با احتمال یک، فرآیند دیریکله یک اندازه احتمال گسسته است.

قضیه ٤.۳. (ستورامن، ۱۹۹٤) فرض کنید :

۱. $V_1, V_2, ..$ متغیرهای تصادفی مستقل و هم‌توزیع از $Beta[1, \alpha]$ و

۲. $X_1, X_2, ..$ متغیرهای تصادفی مستقل و هم‌توزیع با G_0 باشند،

آن‌گاه

$$G \overset{d}{=} \sum_{i=1}^{\infty} W_i \delta_{X_i} \qquad (٤)$$

یک فرآیند دیریکله با پارامترهای (α, G_0) است، که در آن W_iها توابعی نزولی هستند که از فرآیند شکست چوب [۱٤] حاصل می‌شوند. توجه کنید که $E(W_i) = \frac{1}{\alpha+1}(\frac{\alpha}{\alpha+1})^{i-1}$. در واقع $W_1 = V_1$ و برای $i \geq 2$،

$$W_i = V_i \prod_{j=1}^{i-1} (1 - V_j)$$

است.

این قضیه بیان می‌کند اندازه تصادفی G در (؟؟) با احتمال یک، اندازه احتمالی گسسته است، حتی اگر G_0 یک توزیع پیوسته باشد یک نتیجه مهم این قضیه آن است که اگر $G \sim DP(\alpha G_0)$، $X \sim G_0$، $U \sim Beta[1, \alpha]$ و همگی مستقل آن‌گاه $U\delta_X(\cdot) + (1-U)G(\cdot)$ نیز یک فرآیند دیریکله $DP(\alpha G_0)$ است.

[۱٤] Stick-Breaking

٤ مدل‌های مبتنی بر فرآیند دیریکله

١.٤ آمیخته‌ای از فرآیندهای دیریکله

آمیخته‌ای از فرآیند دیریکله[۱۵] (MDP) اولین بار توسط آنتونیاک (۱۹۷٤) معرفی گردید. MDP در مسائل داده‌های سانسورشده و مدل‌های سلسله مراتبی که در آن یک سطح شامل فرآیند دیریکله است کاربرد دارد. به‌طور کلی اگر توزیع پایه G_0 و پارامتر دقت α تصادفی باشد، یعنی

$$G \mid \eta \sim \quad DP(\alpha_\eta G_{0\eta})$$
$$\eta \sim \quad \pi(\eta)$$

آن‌گاه G آمیخته‌ای از فرآیندهای دیریکله با پارامتر دقت α_η، توزیع پایه $G_{0\eta}$ و توزیع آمیختگی π می‌باشد. در این حالت می‌نویسیم

$$G \sim \int DP(\alpha_\eta G_{0\eta}) d\pi(\eta)$$

٢.٤ فرآیند دیریکله آمیخته

اگرچه انتخاب فرآیند دیریکله به عنوان پیشین یک توزیع نامعلوم، مناسب بوده و ویژگی‌های جالبی به همراه دارد، اما هر تحقق از این فرآیند یک توزیع گسسته است و انتخاب آن به عنوان پیشین توزیعی نامعلوم که می‌دانیم پیوسته است موجه به‌نظر نمی‌رسد. یکی از روش‌هایی که برای برطرف کردن این محدودیت فرآیند دیریکله مورد استفاده قرار می‌گیرد، مدل فرآیند دیریکله آمیخته[۱۶] (DPM) است. در این مدل پیچشی[۱۷] از توزیع گسسته G و یک هسته پیوسته که به عنوان توزیع داده‌ها انتخاب می‌شود، مورد استفاده قرار می‌گیرد (فرگوسن، ۱۹۸۳؛ لو، ۱۹۸٤).

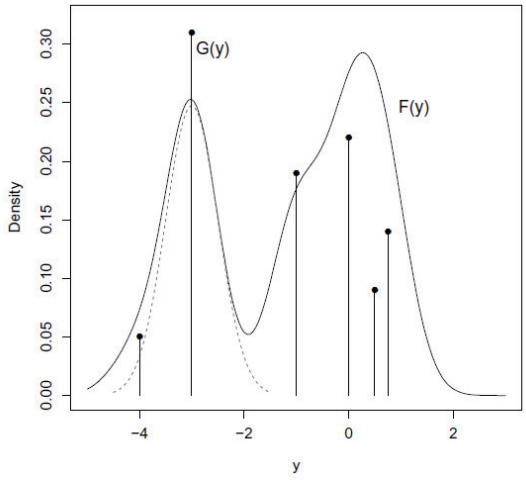
فرض کنید $y_1, y_2, \ldots$ نمونه‌های تصادفی مستقل از توزیع نامعلوم F باشد. پیشین فرآیند دیریکله آمیخته برای F به‌صورت

$$y_i \mid G \sim F(y_i) = \int p(y_i \mid \theta) G(d\theta)$$
$$G \sim DP(\alpha G_0) \qquad (۵)$$

تعریف می‌شود، که در آن $p(y_i \mid \theta)$ یک توزیع پارامتری با پارامتر متناهی بعد θ به عنوان هسته مدل آمیخته است. به‌عنوان مثال در فرآیند دیریکله آمیخته با هسته نرمال مبتنی بر پارامتر میانگین μ داریم

$$y_i \mid G, \sigma^2 \sim \int N(y_i \mid \mu, \sigma^2) G(d\mu)$$

شکل ٤ یک مدل آمیخته با هسته نرمال و $G \sim DP(\alpha G_0)$ را به نمایش می‌گذارد.



شکل ٤. فرآیند دیریکله آمیخته. پیچشی از اندازه احتمال تصادفی G و هسته نرمال.

مدل (؟؟) را می‌توان به صورت سلسه مراتبی

$$y_i \mid \theta_i \overset{iid}{\sim} p(y_i \mid \theta_i)$$
$$\theta_1, \ldots, \theta_n \mid G \overset{iid}{\sim} G$$
$$G \sim DP(\alpha G_0) \qquad (٦)$$

بازنویسی کرد. مدل سلسله مراتبی (؟؟) نیز بر ماهیت خوشه‌بندی θ_i-های تولید شده از اندازه احتمال گسسته G، در یک مدل آمیخته تاکید می‌کند. با حاشیه‌سازی نسبت به G داریم

$$y_i \mid \theta_i \sim p(y_i \mid \theta_i),$$
$$(\theta_1, \ldots, \theta_n) \sim p(\theta_1, \ldots, \theta_n),$$

که در آن توزیع توام $p(\theta_1, ..., \theta_n)$ با توجه به رابطه (؟؟) به دست می‌آید (بلکول، ۱۹۷۳).

مک‌ایچرن و مولر (۱۹۹۸) با قرار دادن نمونه‌های مشابه در یک

[۱۵] Mixture of Dirichlet Processes

[۱۶] Dirichlet Process Mixture Model

[۱۷] Convolution

خوشه و نمونه‌های متفاوت در خوشه‌های مختلف و معرفی مجموعه مقادیر یا خوشه‌های متفاوت با $\theta^* = (\theta_1^*, ..., \theta_{n^*}^*)$ و همچنین $\xi = (\xi_1, ..., \xi_n)$ یک بردار نشانگرهای پیکربندی که $\xi_i = j$ اگر و تنها اگر $\theta_i = \theta_j^*$، نشان دادند

$$y_i \mid \xi_i, \theta_j^* \sim p(y_i \mid \theta_{\xi_i}^*),$$
$$\theta_1^*, \ldots, \theta_{n^*}^* \overset{iid}{\sim} G_0,$$
$$p(\xi_1, \ldots, \xi_n \mid \alpha, n^*) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \alpha^{n^*} \prod_{j=1}^{n^*} \Gamma(n_j),$$

که در آن n^* تعداد خوشه‌ها و $n_j = \sum_i I(\xi_i = j)$ اندازه خوشه $-j$ام است. این تصویر جدید از فرآیند دیریکله آمیخته به‌دلیل خارج کردن توزیع نامتناهی بعدی G و لحاظ نمودن بحث خوشه‌بندی کاربرد بسیاری دارد. خصوصیت خوشه‌بندی فرآیند دیریکله آمیخته سبب شده است که این توزیع آمیخته در مسائلی که خوشه‌بندی داده‌ها مورد نظر است و تعداد خوشه‌ها معلوم نیست به‌طور گسترده مورد استفاده قرار گیرد.

مدل (؟؟) را می‌توان با استفاده از ساختار معرفی شده توسط ستورامن (۱۹۹۴) از فرآیند دیریکله به‌صورت زیر

$$y_i \mid W_h, \theta_h \sim \sum_{h=1}^{\infty} W_h p(y_i \mid \theta_h) \qquad (۷)$$

نیز بازنویسی کرد که در آن

$$\theta_h \overset{iid}{\sim} G_0$$
$$W_h = V_h \prod_{k<h} (1 - V_h)$$
$$V_h \overset{iid}{\sim} Beta(1, \alpha)$$
$$G = \sum_{h=1}^{\infty} W_h \delta_{\theta_h}$$

مدل (؟؟) به این نکته اشاره دارد که یک مدل فرآیند دیریکله آمیخته را می‌توان به‌صورت آمیخته گسسته‌ای از توزیع‌های $p(y_i \mid \theta_h)$ با وزن‌های W_h در نظر گرفت. یکی از نتایج این مدل آن است که مدل DPM بین مشاهده علمی یک خوشه‌بندی ایجاد می‌کند و α تعداد این خوشه‌ها را کنترل می‌کند. اگر $\alpha \to 0$ یعنی تعداد خوشه‌ها کاهش می‌یابد به این معنی که تمامی θ_i–ها رفتاری مشابه دارند و مدل به یک مدل پارامتری تبدیل می‌شود، که $\theta \sim G_0$ و داده‌ها به‌طور تصادفی و مستقل از $p(y \mid \theta)$ تولید می‌شوند. در مقابل هنگامی که $\alpha \to \infty$ یعنی تعداد خوشه‌ها بسیار زیاد می‌شود به این معنی که هر θ_i در یک خوشه قرار می‌گیرد و رفتاری مستقل از بقیه θ–ها دارد و بنابراین

$$y_i \overset{iid}{\sim} \int p(y_i \mid \theta) G_0(d\theta).$$

۳.۴ فرآیند دیریکله متناهی

با توجه به ساختار معرفی شده توسط ستورامن (۱۹۹۴) از فرآیند دیریکله نتیجه می‌شود که شبیه‌سازی یک اندازه احتمال تصادفی G نیازمند شبیه‌سازی تعداد نامتناهی از متغیرهای تصادفی است. این موضوع منجر به پیچیدگی‌های محاسباتی خواهد شد و عملکرد روش بیز ناپارامتری را تحت تاثیر قرار می‌دهد. به‌علاوه با توجه به این‌که وزن‌ها دنباله‌ای نزولی تشکیل می‌دهند در نظر گرفتن مقدار نامتناهی از متغیرهای تصادفی غیرضرور می‌نماید. با توجه به این مسائل، مولیر و تاردلا (۱۹۹۸) فرآیند دیریکله بریده شده[۱۸] را تحت عنوان فرآیند ϵ–دیریکله به‌صورت زیر تعریف کردند:

تعریف ۱.۴. برای هر $\epsilon \in (0, 1)$ توزیع تصادفی G^ϵ یک فرآیند ϵ–دیریکله است اگر

$$G^\epsilon(\cdot) = \sum_{h=1}^{H_\epsilon} W_h \delta_{\theta_h}(\cdot) + \{1 - \sum_{h=1}^{H_\epsilon} W_h\} \delta_{\theta_{H_\epsilon+1}}(\cdot). \qquad (۸)$$

که در آن θ_h به‌طور تصادفی و مستقل از توزیع پایه G_0 نمونه‌گیری می‌شود،

$$V_h \overset{iid}{\sim} Beta(1, \alpha) \text{ ،} W_h = V_h \prod_{k<h}(1 - V_k)$$

و

$$H_\epsilon = \inf\{m \in Nn : \sum_{h=1}^{m} W_h \geq 1 - \epsilon\}$$

است. ساختار تعریف شده در (؟؟) همانند ساختار فرآیند دیریکله است. اما این ساختار پس از تعداد تصادفی از جملات متوقف می‌شود و جرم احتمال باقیمانده به نقطه‌ای تصادفی در χ که به‌طور مستقل از G_0 تولید شده، اختصاص می‌یابد. در تعریف فرآیند ϵ–دیریکله نقش زمان توقف، یعنی H_ϵ این است که اجازه دهد احتمال تصادفی تا حد مورد نظر نزدیک به فرآیند دیریکله تولید گردد. مولیر و تاردلا (۱۹۹۸) نشان دادند $H_\epsilon \sim Poi(-\alpha \log \epsilon)$ که Poi نشان دهنده توزیع پواسن است. همچنین نشان دادند که

[۱۸] Truncated Dirichlet Process

اگر

$$G(\cdot)=\sum_{h=1}^{\infty}W_h\delta_{\theta_h}$$
$$G^{\epsilon}(\cdot)=\sum_{h=1}^{H_\epsilon}W_h\delta_{\theta_h}(\cdot)\{1-\sum_{h=1}^{H_\epsilon}W_h\}\delta_{\theta_{H_\epsilon+1}}(\cdot),$$

آن‌گاه

$$\sup_B\{|G(B)-G^{\epsilon}(B)|\}\leq\epsilon.$$

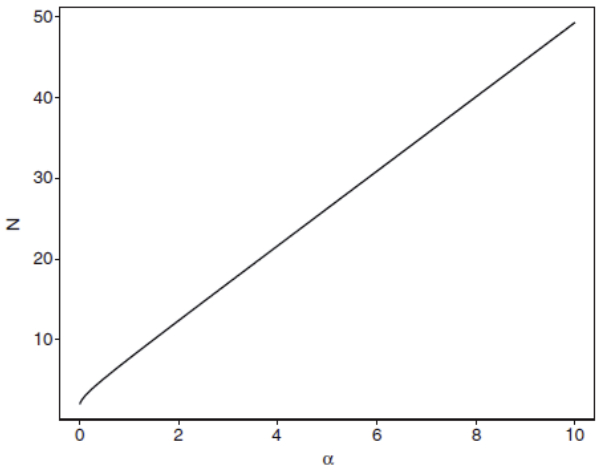
در واقع هنگامی $\epsilon\to 0$ نمونه‌های تولید شده از فرآیند ϵ-دیریکله به فرآیند دیریکله همگرا می‌شوند. به‌عبارت دیگر با در نظر گرفتن یک مقدار $\epsilon>0$ مناسب، می‌توان تقریبی مناسب از فرآیند دیریکله در اختیار داشت. لازم به ذکر است ایشواران و جیمز (۲۰۰۱) تعریف دیگری از فرآیند دیریکله بریده شده ارائه کردند. در این تعریف تقریبی دیگر از فرآیند دیریکله به‌صورت

$$\sum_{i=1}^{\infty}W_i\delta_{\theta_i}\approx\sum_{i=1}^{N}W_i\delta_{\theta_i}$$

به عنوان فرآیند دیریکله بریده شده معرفی شد و نشان دادند $G\sim TDP(\alpha G_0,N)$. در این تقریب $W_N=1-\sum_{i=1}^{N-1}W_i$ بوده و پارامتر N تعداد نقاط جرم مورد استفاده را مشخص می‌کند. همچنین مقدار N رابطه مستقیم با α دارد که کنترل کننده خوشه‌ها در بین واحدهای آزمایشی است. هنگامی که N کوچک باشد محاسبات ساده است اما تقریب مورد نظر دقیق نیست. رویکردی که برای انتخاب N پیشنهاد گردید بر این اساس بود که مقدار N به‌گونه‌ای انتخاب شود که مقدار مورد انتظار W_N بسیار کوچک باشد به‌عبارت دیگر $E(W_N)\approx\epsilon$

ایشواران و جیمز (۲۰۰۱) نشان دادند در این صورت داریم

$$N\approx 1+\frac{\log\epsilon}{\log(\frac{\alpha}{1+\alpha})}.$$



شکل ۵. نمودار مقدار N در برابر α هنگامی که $E(\sum_{i=1}^{N-1}W_i)>0/99$.

شکل ۵ نمودار α در برابر N هنگامی که $\epsilon=0/01$ انتخاب شده را به نمایش می‌گذارد. همان‌طور که مشاهده می‌شود رابطه‌ای تقریباً خطی بین دو پارامتر برقرار است.

۵ فرآیند دیریکله در ساختار مدل‌های متغیر پنهان

متغیر پنهان، در مقابل متغیرهای آشکار، متغیرهایی هستند که به صورت مستقیم مشاهده نمی‌شود، اما به واسطه متغیرهای دیگر که قابل مشاهده هستند توسط یک مدل آماری استنباط می‌شوند. به عبارت دیگر متغیر پنهان، متغیری است که به طور مستقیم مشاهده نمی‌شود اما منبعی است که موجب تغییرات پنهان در متغیرهای آشکار مدل یا پارامترهای مدل می‌شود. به طور مثال فرض کنید هدف، مطالعه و بررسی عملکرد یک دارو بر روی تعدادی از کودکان بیمار در یک جامعه باشد. در این جامعه علاوه بر متغیرهای آشکاری مانند سن، جنسیت و وزن که قابل اندازه‌گیری هستند یک مجموعه از عوامل ژنتیکی در نتیجه آزمایش موثر است که به عنوان متغیر پنهان می‌توان در مدل در نظر گرفت. عموماً فرض می‌شود توزیع پیشین متغیرهای پنهان نرمال است. این در حالی است که به دلیل عدم مشاهده این متغیرها صحت چنین فرضی سؤال برانگیز است.

برای حل این مسأله، می‌توان از فرآیند دیریکله به عنوان یک رهیافت بیزی ناپارامتری برای مدل‌بندی متغیرهای پنهان استفاده

کرد. مدل سلسله مراتبی

$$\begin{aligned} Y_i|b_i,\eta &\overset{ind}{\sim} f(y_i|b_i,\eta), \\ b_i|G &\overset{iid}{\sim} G, \\ G &\sim DP(\alpha,G_0) \end{aligned} \qquad (۹)$$

را در نظر بگیرید. در مدل (؟؟) Y_i به شرط پارامتر η و متغیر پنهان b_i از خانواده توزیع پارامتری f آمده‌اند. η–ها پارامترهای مدل بوده و b_i–ها متغیرهای پنهان هستند که از توزیع نامعلوم G تولید شده‌اند. G تحققی از یک فرآیند دیریکله با پارامترهای α و G_0 است. بلکول و مک‌کویین (۱۹۷۳) نشان دادند که با حاشیه‌سازی و انتگرال‌گیری روی G در مدل (؟؟)، b_i–ها از توزیع‌های شرطی زیر

$$\begin{aligned} b_i|b_1,\cdots,b_{i-1},\alpha,G_0 \\ \sim \tfrac{1}{\alpha+i-1}\textstyle\sum_{h=1}^{i-1}\delta_{b_h} \\ +\tfrac{\alpha}{\alpha+i-1}G_0 \end{aligned} \qquad (۱۰)$$

پیروی می‌کنند. رابطه(؟؟) در واقع بیانگر این است که اثرات تصادفی با احتمال $\frac{1}{\alpha+i-1}$ مشابه یکی از مقادیر قبلی بوده و با احتمال $\frac{\alpha}{\alpha+i-1}$ یک مقدار جدید از G_0 را می‌گیرد. همان‌طور که گفتیم استفاده از این ساختار به دلیل عدم وجود توزیع با بعد نامتناهی G، سادگی محاسبات را به دنبال دارد. اما مفهوم خوشه‌بندی به عنوان یک ویژگی مهم DP را شکل نمی‌دهد. بنابراین مشابه آنچه که در بخش‌های قبل بیان شد با استفاده از ویژگی خوشه‌بندی DP می‌توان یک مدل ساده‌تر و کاراتر را ارائه داد. یکی از کاربردهای مدل متغیر پنهان در تحلیل داده‌های دو–دویی است. در این چارچوب، داده‌های دو–دویی را به یک متغیر پنهان پیوسته ارتباط می‌دهند به این صورت که با استفاده از یک مقدار آستانه متغیر پیوسته در دو رده گسسته شده و متغیر پاسخ معرفی می‌شود. به‌طور کلی در اغلب متون، تحقیقات به عمل آمده فرض می‌شود توزیع پیشین متغیر پنهان نرمال است و بدین ترتیب مدل حاصل پروبیت[۱۹] نامیده می‌شود. این در حالی است که چنین فرضی ممکن است برقرار نباشد. برای حل این مسأله رهیافت‌های متفاوتی وجود دارد (کافو و همکاران، ۲۰۰۷). یک رهیافت بر مبنای اتخاذ رهیافت بیز نیمه‌پارامتری[۲۰] است که در آن پارامترهای مدل متغیر پنهان به دو دسته تقسیم شده و برای برخی از آنها روش پارامتری و برای بقیه چارچوبی ناپارامتری بر اساس فرآیند دیریکله فرض می‌شود. بدین ترتیب یک مدل پروبیت تعمیم‌یافته با متغیر پنهان از فرآیند دیریکله آمیخته[۲۱] شکل می‌گیرد (جارا و همکاران، ۲۰۰۷). در مدل‌های آمیخته خطی تعمیم‌یافته[۲۲] نیز برای مدل‌بندی اثرات تصادفی اغلب از توزیع نرمال استفاده می‌شود. اما ممکن است این فرض برقرار نباشد. از این رو استفاده از پیشین فرآیند دیریکله برای توزیع نامعلوم اثرات تصادفی منجر به استنباط دقیق‌تر و کارایی بیشتر مدل می‌شود. کیونگ و همکاران (۲۰۱۱) برای مدل‌بندی اثرات تصادفی از فرایند دیریکله استفاده کردند.

۶ بحث و نتیجه‌گیری

در این مقاله فرآیند دیریکله و ویژگی‌های مهم آن به عنوان یک پیشین مناسب در مدل‌های بیز ناپارامتری معرفی شد. همان‌طور که بیان شد این فرآیند ضمن این‌که به عنوان یک پیشین مزدوج در تسریع محاسبات بیزی نقش مهمی دارد با ویژگی خوشه‌بندی داده‌ها ناهمگنی بین آن‌ها را کنترل می‌کند و این امر منجر به بهبود و کارایی مدل می‌شود. این خواص DP در سال‌های اخیر موجب محبوبیت آن در گستره‌ی وسیعی از مسائل پزشکی مانند پردازش تصاویر و اپیدمیولوژی شده است. یکی از کاربردهایی که می‌توان به عنوان تحقیقات آینده مد نظر داشت استفاده از این فرآیند در مدل‌بندی بیزی ناپارامتری داده‌های فضایی است.

[۱۹] Probit Model

[۲۰] Bayesian Semiparametric Model

[۲۱] Dirichlet Process Mixture

[۲۲] Generalized Linear Mixed Models

مراجع

[1] Antoniak, C. E. (1974). Mixture of Dirichlet Processes with Applications to Bayesian Nonparametric Problem, *The Annals of Statistics*, **2**, 1152-1174.

[2] Blackwell, D. (1973). Discreteness of Ferguson Selection, *The Annals of Statistics*, **1**, 356-358.

[3] Blackwell, D. and MacQueen, J. (1973). Ferguson Distributions Via Polya urn Schemes, *The Annals of Statistics*, **1**, 353-355.

[4] Caffo, B., An, M. and Rohde, C. (2007). Flexible Random Intercept Models for Binary Outcomes Using Mixtures of Normal, *Journal of Computational Statistics and Data Analysis*, **51**, 5220-5235.

[5] Dey, D., Muller, P. and Sinha, D. (1998). Practical Nonparametric and Semiparametric Bayesian Statistics, New York, *Springer-Verlag*.

[6] Malcolm, F. (2008). Bayesian Statistics, *Newcastle University*.

[7] Ferguson, T. S. (1973). A Baysian Analysis of Some Non-Parametric Problems, *Annals of Statistics*, **1**, 209-230.

[8] Ferguson, T. S. (1983). Bayesian Density Estimation by Mixtures of Normal Distributions. In Recent Advances in Statistics, eds. J. Rustage and G. G. Rizvi, *New York: Academic Press*, 287-302.

[9] Ghosh, J. K. and Ramamoorthi, R. V. (2003). Bayesian Nonparametric, *New York:Springer Series in Statistics* .

[10] Ishwaran, H and James, L. F. (2001), Gibbs Sampling Methods for Stick-Breaking Priors, *Journal of American Statistical Association*, **96**, 161-173.

[11] Jara, A. Garc´Ia-Zattera, M. J. and Lesaffere, E. (2007). A Dirichlet Process Mixture Model for the Analysis of Correlated Binary Responses. *Computational Statistics and Data Analysis*, **51**, 5402-5415.

[12] Kyung, M. Gill, J. and casella, G. (2011). Sampling Schemes for Generalized Linear Dirichlet Process Random Effects Model, *Journal of Statistical Methods and Applications*, **20**, 303-304.

[13] Kolmogorov, A. N. (1933), Foundations of the Theory of Probability, 2nd Ed. *New York, USA: Nathan Morrison* .

[14] Lo, A. Y. (1984), On a Class of Bayesian Nonparametric Estimates: Density Estimates, *The Annals of Statistics*, **12**, 351-357.

[15] MacEachern, S. N. and Muller, P. (1998). Estimating Mixture of Dirichlet Process Models, *Journal of Computation and Grafical Statistics*, **2**, 223-238.

[16] Muliere, P. and Tardella, L. (1998). Aproximating Distributions of Functionals of Ferguson-Dirichlet priors, *Canadian Journal of Statistics*, **30**, 269-283.

[17] Muller, P. and Rodriguez, A. (2013). *Nonparametric Bayesian Inference*, IMS-CBMS Lecture Notes. IMS.

[18] Muller, P. and Mitra, R. (2013). Bayesian Nonparametric Inference -Why and How, *Baysian Analysis*, **8**, 1-34.

[19] Sethuraman, J. (1994). A Constructive Definition of Dirichlet Prior, *Statistica sinica*, **2**, 639-650.

[20] Tsiatis, A. A. (2006). Semiparametric Theory and Missing Data, *New York USA:Springer*.