مروری بر استنباط نیرومند بیزی

علیرضا دانشخواه*　　　سیامک نوربلوچی†

چکیده

مفهوم استنباط نیرومند بیزی[1] و سه معیار مختلف سنجش استنباطها معرفی می‌شود. با خانوادهٔ پیشینهای رایج در مطالعات نیرومندی به ویژه رده εـ آلوده‌ها و چند فن یافتن دامنهٔ تغییرات معیاری پسینی آشنا می‌شویم.

مقدمه

از جنبه‌ای، روشهای استنباط آماری را می‌توان به دو دستهٔ روشهای استنباط کلاسیک و روشهای استنباط بیزی تقسیم کرد. در استفاده از روشهای استنباط بیزی برای مسأله‌ای خاص، علاوه بر تعیین توزیع احتمال نمونه، براساس اطلاعات پیشین مرتبط با مسألهٔ موردنظر، باید توزیعی را به عنوان توزیع پیشین برای این اطلاعات پیشین مسألهٔ مذکور ارائه دهیم.

اما در عمل، مدلبندی اطلاعات پیشین در قالب یک توزیع پیشین به دلیل عدم حتمیتی که نسبت به آن وجود دارد، مهمترین مشکل در این روشهای استنباط آماری است.

علاوه بر مشکل فوق، اختلاف‌نظر بین آمارشناسان منجر به کارگیری توزیعهای پیشین متفاوتی می‌گردد و چنین اختلافی شاید منجر به استنباطهای متفاوتی شود. در واقع، استنباطی که بر اساس یک توزیع پیشین بهینه است، ممکن است تحت توزیعهای پیشین دیگر بهینه نباشد.

آنچه در بالا آمد، انگیزه اساسی طرح مبحث نیرومندی بیزی است که منشأ آن را در مجموعهٔ مقالات گود[2] در بین سالهای ۱۹۵۰ تا ۱۹۸۳ می‌توان مشاهده کرد. در واقع گود مبحث نیرومندی بیزی را معرفی کرد و سپس در سالهای اخیر دیگران آن را گسترش داده‌اند. دیدگاهها و اندیشه‌های گود را به طور خلاصه در بخش دوم معرفی می‌کنیم.

مقالهٔ حاضر مروری بر فعالیتهای اخیر در این زمینه است (مرور کلی و جامعی را برگر[3] (۱۹۸۴) ارائه کرده است). عمده‌ترین فعالیتهایی که اخیراً در این حوزه انجام شده عبارت‌اند از:

۱) مدلبندی اطلاعات پیشین در یک ردهٔ معین Γ متشکل از توزیعهای پیشین و ممکن.

۲) تعیین دامنهٔ تغییرات معیاری پسینی، برای انواع گوناگونی از Γها. وقتی توزیع پیشین در ردهٔ پیشینهای Γ، تغییر می‌کند.

در بخش سوم به معرفی و بررسی رده‌های گوناگون توزیعهای پیشین می‌پردازیم و در بخش چهارم به معرفی روشهایی برای تعیین دامنهٔ تغییرات معیارهای پسین، وقتی که توزیع پیشین در ردهٔ εـ آلوده‌ها تغییر می‌کند، می‌پردازیم. بخش پنجم را به مقایسه‌ای بین نتایج به دست آمده از رده‌های مختلف آلودگی در ردهٔ توزیع‌های پیشین εـ آلوده اختصاص می‌دهیم. قبل از ورود به بحث، معرفی چند نماد مفید و ضروری به نظر می‌رسد.

*) علیرضا دانشخواه، گروه آمار دانشگاه شهید چمران ـ اهواز　†) سیامک نوربلوچی، گروه آمار دانشگاه شهید بهشتی ـ تهران

1) Bayesian Robustness　2) Good, I. J　3) Berger, J. O

X، بیانگر متغیر (بردار) تصادفی قابل مشاهده‌ای است که فرض می‌شود دارای تابع چگالی $f(x|\theta)$ (نسبت به اندازهٔ لبگ) است، θ پارامتری نامعلوم است که در فضای پارامتر Θ تغییر می‌کند. پیشین تعریف شده بر Θ را با π نمایش می‌دهیم و چگالی حاشیه‌ای نسبت به این پیشین را با

$$m(x|\pi) = E^{\pi} f(x|\theta) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta \qquad (۱)$$

نمایش می‌دهیم.

چگالی پسین θ به شرط $X = x$ (با شرط وجود) نسبت به پیشین π را با $\pi(.|x)$ نشان می‌دهیم که به صورت زیر تعریف می‌شود:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x|\pi)}, \qquad (۲)$$

و در پایان Q بیانگر فضای تمامی توزیعهای احتمال تعریف شده بر Θ است.

۱ فلسفهٔ گود و معرفی آنالیز بیزی نیرومند

علی‌رغم مزایایی که روشهای استنباط بیزی دارند، حالتهای بسیاری هست که نمی‌توان اطلاعات قبلی راجع به پارامتر را توسط توزیع پیشین واحدی مدلسازی کرد. در چنین مسائلی مطالعه و به‌کارگیری روشهای استنباط بیزی نیرومند توصیه می‌شود. قبل از آشنایی با این روشها و معرفی فلسفه گود در این زمینه، بهتر است خواننده را با ذکر مثالی با موضوع آشنا سازیم.

مثال ۱: فرض می‌کنیم $X \sim N(\theta, ۱)$ و θ نامعلوم است. می‌دانیم میانه، چارک اول، و چارک سوم توزیع پیشین به ترتیب برابر ۰ ، ۱− و ۱ هستند. با توجه به این اطلاعات می‌توان توزیع کوشی $\pi_c = C(۰, ۱)$ یا توزیع نرمال $\pi_N = N(۰, ۲٫۱۹)$ را به عنوان توزیع پیشین در نظر گرفت. اما سؤالی که بلافاصله به ذهن می‌آید، این است آیا از بین π_c و π_N یکی بر یکدیگری ارجحیت دارد یا خیر، به عبارت دیگر آیا نتیجهٔ استنباط وابستگی به انتخاب π_c یا π_N دارد یا نه.

طبیعی‌ترین راه جهت پاسخ به سؤال فوق بررسی کردن این مطلب است که آیا واقعاً نتایج استنباط بر اساس این انتخاب تغییر می‌کنند یا نه. با درنظر گرفتن توان دوم خطاها به عنوان تابع زیان θ، برآوردهای بیزی θ نسبت به دو توزیع پیشین π_c و π_N را به ترتیب با δ^cوδ^N نشان می‌دهیم. جدول زیر بیانگر مقادیر مختلف این برآوردها به ازای چند مشاهده است:

x	۰	۱	۲	۴٫۵	۱۰
$\delta^C(x)$	۰	۰٫۵۲	۱٫۲۷	۴٫۰۹	۹٫۵
$\delta^N(x)$	۰	۰٫۶۹	۱٫۳۷	۳٫۰۹	۶٫۸۷

همان طور که از جدول بالا معلوم است، برای مقادیر کوچک x، $(x \leq ۲)$، $\delta^C(x)$ و $\delta^N(x)$ تقریباً نزدیک به یکدیگرند، پس می‌توان نتیجه گرفت که برای مقادیر کوچک x فرقی نمی‌کند، از $\pi_c(x)$ یا $\pi_N(x)$ استفاده شود. استنباطی که به طور جدی تحت تأثیر تغییرات فرضهای مطروحه در آن نباشد «نیرومند» نامیده شده است. استنباط فوق برای xهای کوچک نسبت به انتخاب پیشین نیرومند است.

اکنون که با شرح مثال فوق، تا حدودی با مفهوم نیرومندی نسبت به تعیین نادرست توزیع پیشین آشنا شده‌ایم، مبانی پیدایش این اندیشه را شرح می‌دهیم.

گود اولین آمارشناس بیزی است که تعیین نادرست توزیع پیشین را به عنوان بخش عمده‌ای از فلسفهٔ آماری خود مطرح کرده است. البته باید توجه کرد توصیفهای گود از دیدگاه بیزی نیرومند از نظر فلسفی تا به حال گسترش چندانی نیافته است. اکثر آمارشناسان بیزی به آنها معتقد بوده و روشهایی را برای استفاده از این دیدگاه در مسائل مختلف ارائه کرده‌اند. گود (۱۹۷۳) بیست و هفت اصل آماری را برای فلسفهٔ آماری خود ارائه کرده است که برخی از آنها به طور ضمنی بر روش بسیار مفید رایج در ارزیابی بیزی نیرومند یعنی روش پسین نیرومندی دلالت می‌کنند. وی برای تضمین اعتبار این روش، چند اصل را که در واقع بر پسین نیرومندی دلالت می‌کنند در مدلی که موسوم به مدل «جعبه سیاه»[1] است، پیشنهاد کرده است. در نظر گرفتن جعبه‌ای که شامل تمام روشهای معمول بیزی است، ایدهٔ اصلی مدل مذکور است. این جعبه به عنوان خروجی انواع استنباطهای آماری را تولید می‌کند. یکی از مهمترین این روشهای پیشنهادی، به نام روش ارزیابی پسین نیرومندی را شرح می‌دهیم. فرض کنید می‌خواهیم کمیت پسین مورد نظر $\rho(x,\pi)$ (به طور مثال، میانگین پسین، واریانس پسین، احتمال پسین، ناحیهٔ مورد اعتماد پسین فرض H، مقدار مورد انتظار پسین) را مورد ارزیابی قرار دهیم و وقتی که توزیع پیشین π در رده‌ای مانند Γ تغییر می‌کند، تغییرات این کمیت را مورد بررسی قرار دهیم. به عبارتی می‌خواهیم حدود تغییرات زیر را تعیین کنیم،

$$\underline{\rho}(x,\pi) = \inf_{\pi\in\Gamma} \rho(x,\pi), \bar{\rho}(x,\pi) = \sup_{\pi\in\Gamma} \rho(x,\pi). \qquad (۳)$$

البته امیدواریم که طول بازهٔ $(\underline{\rho}(x,\pi), \bar{\rho}(x,\pi))$ تحت هر یک از توزیعهای پیشین در ردهٔ Γ به اندازهٔ کافی کوچک باشد به طوری که بتوان تأثیر اختلاف بین پیشینهای موجود در ردهٔ Γ را بر استنباط بیزی (متکی بر یکی از پیشینهای متعلق به Γ) ناچیز دانست و با اطمینان کافی از آن استفاده کرد. البته قضاوت در مورد میزان کمی یا زیادی طول بازهٔ مذکور بستگی به مسأله مورد مطالعه دارد و نمی‌توان حکمی کلی در مورد آن ارائه کرد. رهیافتهای

1) Black Box

گوناگونی برای ارزیابی نیرومندی معرفی شده‌اند که برخی از آغاز متکی بر دیدگاه بیزی و برخی دیگر متکی بر دیدگاه فراوانی (در دراز مدت) هستند. روش پسین نیرومندی، یکی از معروفترین و مهمترین این رهیافتها است که در اغلب مسائل می‌توان آن را به کار برد. برای آگاهی از مزیتها و معایب این روش و همچنین آگاهی از سایر رهیافتها، برگر (۱۹۸۴) را مطالعه کنید. به همین منظور در مقالهٔ حاضر تحلیل نیرومندی بیزی بر اساس روش پسین نیرومندی صورت گرفته است. در سالهای اخیر مقالات متعددی بر این اساس نوشته شده است.

۲ دامنهٔ تغییرات معیارهای بیزی

۱.۲ معیارهای مورد نظر

سه دسته از معیارهای بیزی به طور شاخص مورد نظر قرارگرفته‌اند. دسته‌های تعیین شده به شکل توزیع پیشین بستگی دارند.

۲.۲ تابعکهای خطی

آسانترین معیار موجود، از نظر دیدگاه بیزی نیرومند، تابعکهای خطی در π هستند که صورت انتگرالی زیر را دارند:

$$Q(\pi) = \int_{\Theta} h(\theta)\pi(\theta)d\theta \qquad (۴)$$

که در آن h تابعی معین است. شاید بتوان گفت، یکی از مهمترین تابعکهای خطی به صورت زیر است:

$$m(x|\pi) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta, \qquad (۵)$$

یعنی چگالی حاشیه‌ای X نسبت به توزیع پیشین π. بر اساس این نمادگذاری، توزیع پیشین ML-II[1] تعریف می‌شود. فرض کنید Γ رده‌ای از توزیعهای پیشین است، در این صورت اگر $\hat{\pi} \in \Gamma$ در رابطهٔ $m(x|\hat{\pi}) = \sup_{\pi\in\Gamma} m(x|\pi)$ صدق کند، $\hat{\pi}$ پیشین درستنمایی ماکسیمم نوع دوم یا پیشین ML-II نامیده می‌شود.

۳.۲ نسبت معیارهای پسین خطی

بسیاری از معیارهای پسین مورد علاقه را می‌توان به صورت زیر نوشت:

$$\rho(x|\pi) = \int_{\Theta} h(\theta)f(x|\theta)\pi(\theta)d\theta/m(x|\pi). \qquad (۶)$$

برای تابعی مانند h به طور مثال با انتخاب $h(\theta) = \theta$، میانگین پسین به دست می‌آید. همچنین با انتخاب $h(\theta) = I_C(\theta)$ (تابع نشانگر مجموعهٔ C) احتمال پسین مجموعهٔ C نتیجه می‌شود. دلیل نامگذاری نسبت معیارهای پسین خطی، توانایی نوشتن این معیارها به صورت نسبتی از تابعکهای خطی در π است.

۴.۲ نسبت معیارهای ناخطی پسین

معیارهای پسینی هستند به صورت زیر

$$\rho(x|\pi) = \int_{\Theta} h(\theta,\varphi(\pi))f(x|\theta)d\theta/m(x\pi) \qquad (۷)$$

که در آن h تابعی غیرخطی است.

مهمترین مثال، آن واریانس پسین است که توسط تابعک زیر تعریف می‌شود:

$$V(X|\pi) = \int_{\Theta} (\theta-\mu(x|\pi))^{۲}\pi(\theta)d\theta/m(x|\pi) \qquad (۸)$$

که در آن $\mu(x|\pi)$ میانگین پسین است.

۵.۲ رده‌های پسین

همان گونه که اشاره کردیم، هرگز نمی‌توان باورهای پیشین را بدون خطا، کمّی کرد. در این وضعیت، پس از پایان فرآیند استخراج توزیع پیشین، آمارشناس بیزی با خانوادهٔ Γ از توزیعهای پیشین که منعکس کنندهٔ باورهای پیشین او است رو به رو خواهد شد. این خانواده به گونه‌ای است که π_Γ (توزیع پیشین واقعی) عضوی ناشناخته از آن است.

در انتخاب Γ باید چهار هدف عمده زیر تأمین شوند:

۱) Γ باید به طوری انتخاب شود که ارزیابی نیرومندی نسبت به آن، به سادگی میسر باشد.

۲) برای آنکه اعتماد کافی به نیرومندی استنباط (به عمل آمده) داشته باشیم، Γ باید تا حد امکان شامل توزیعهای پیشین معقول (برای مسأله) باشد.

۳) نباید شامل پیشینهای نامعقولی باشد که به غلط نتیجه بگیریم، استنباط حاصله نیرومند نیست.

۴) Γ باید متناظر با اطلاعات پیشینی باشد که به آسانی قابل استخراج هستند.

1) Maximum likelhood-type II prior

با توجه به اهداف فوق یا مصالحه‌ای از آنها، صورتهای مختلفی برای Γ به ترتیب زیر پیشنهاد می‌شود.

۶.۲ ردهٔ توزیعهای پیشین مزدوج[1]

تعریف (۱ـ۲ـ۳) گیریم π_λ یک پیشین مزدوج متناظر با تابع درستنمایی $L(\theta)$ باشد. ردهٔ پیشنهادی مزدوج را که با Γ_C نشان خواهیم داد، به صورت زیر تعریف می‌کنیم:

$$\Gamma_C = \{\pi_\lambda : \lambda \in \Lambda\},$$

که در آن Λ مجموعهٔ مقادیر معقولی از ابر پارامتر λ است (برگر (۱۹۸۵) را ببینید). این رده، شامل مزایا و معایبی است. مهمترین مزایای این رده، محاسبهٔ آسان معیارهای پسین برای بسیاری از توابع پارامتری و ارزیابی آنها و سادگی به دست آوردن اکثر معیارهای ارزیابی نیرومندی است. با وجود این مزایا، این رده معایبی نیز دارد. از جمله اینکه ممکن است بسیاری از توزیعهای معقول را شامل نگردد. از طرف دیگر، به طور کلی دمهای توزیعهای پیشین موجود دراین رده از تغییرات زیادی برخوردار نیستند. (به طورکلی، رده‌هایی ارجح هستند که به قدرکافی بزرگ بوده و شامل پیشینهای معقول برای مسألهٔ مورد مطالعه باشند یا لااقل پیشینهایی با صورتهای تابعی مختلف را در بر بگیرند).

۷.۲ ردهٔ گشتاوری[2]

تعریف (۲ـ۲ـ۳) ردهٔ گشتاوری را که با Γ_M نشان می‌دهیم به صورت زیر تعریف می‌کنیم:

$$\Gamma_M = \{\pi : \alpha_i \leq E^\pi(\theta^i) \leq \beta_i, i = ۱, ۲, \ldots, K\}.$$

معمولترین ردهٔ گشتاوری که به کار می‌رود، رده‌ای است که بر حسب دو گشتاور اول و دوم تعریف می‌شود. برای بحث بیشتر در این باره به استون[3] (۱۹۶۳)، هارتیگان[4] (۱۹۶۹)، گلدشتاین[5] (۱۹۸۰) مراجعه کنید. معایب اصلی Γ_M نیز مانند Γ_C است. هر دو رده محدودیتهای شدیدی بر شکل دم توزیع پیشین و توانایی مدلبندی آن اعمال می‌کنند.

۸.۲ ردهٔ نسبت چگالیها[5]

فرض کنید به ازای هر پیشین (متعلق به ردهٔ Γ) اندازه غالب واحدی مانند ν وجود دارد به طوری که نسبت به این اندازه، چگالیهای پیشین قابل تعریف‌اند. در این صورت، ردهٔ نسبت چگالیها بر اساس این پیشینها، به صورت زیر تعریف می‌شود.

تعریف (۳ـ۲ـ۳) فرض کنید L و U تابعهایی نا منفی و مشخص باشند. ردهٔ چگالیهای نسبتی را که با نماد Γ_{Dr} نشان می‌دهیم، به صورت زیر تعریف می‌کنیم:

$$\Gamma_{Dr} = \{\pi : L(\theta) \leq \alpha\pi(\theta) \leq U(\theta), \text{به ازای } \alpha \text{ یی}\}.$$

تعریف دیگر این رده که از نظر عملی کاربرد بیشتری دارد، به صورت زیر است:

$$\Gamma_{Dr} = \{\pi : \frac{L(\theta)}{U(\theta')} \leq \frac{\pi(\theta)}{\pi(\theta')} \leq \frac{V(\theta)}{L(\theta')}, \text{برای تمام } \theta \text{ها و } \theta' \text{ها}\}$$

از آنجا که معمولاً توزیع پیشین را از مقایسه دو به دوی نقاط فضای پارامتر (بر اساس یک معیار مطلوبیت) تعریف می‌کنند، تعریف بالا معقول و عملی به نظر می‌رسد. جهت آگاهی بیشتر در این باره و مزایا و معایب این رده به مقاله درو برتیس[6] و هارتیگان (۱۹۸۱) مراجعه کنید.

از دیگر رده‌های مهم، می‌توان از رده‌های چندکی، توزیعهای با چگالی کراندار، εـ آلوده‌ها نام برد که فقط به بررسی ردهٔ اخیر می‌پردازیم و برای آگاهی در مورد دو ردهٔ اول برگر (۱۹۹۰) را ببینید.

۳ نیرومندی بیزی تحت ردهٔ εـ آلوده‌ها

۱.۳ ردهٔ εـ آلوده‌ها

در عمل، پس از فرایند استخراج توزیع پیشین، به توزیعی مانند $\pi_۰$ که بیانگر تقریبی باورها و اطلاعات پیشین است، می‌رسیم. یک روش معقول جهت بیان عدم یقین کامل به درستی $\pi_۰$، در نظرگرفتن پیشینهای مجاور $\pi_۰$ است. به منظور تعریف مجاورت دو توزیع، معیارهای گوناگونی برای اندازه‌گیری فاصله بین دو توزیع به‌کار می‌روند. یکی از این معیارها، معیاری است که در تعریف ردهٔ پیشینهای ε ـ آلوده‌ها به کار رفته است. ردهٔ مذکور به صورت زیر است:

$$\Gamma_\varepsilon = \{\pi = (۱ - \varepsilon)\pi_۰ + \varepsilon q : q \in Q\} \qquad (۹)$$

که در آن $\varepsilon \in [۰, ۱]$ داده شده و بیانگر میزان عدم حتمیت در مورد $\pi_۰$ و Q

1) class of conjugate priors 2) moment class 3) M. Stone 4) A. Hartigan 5) M. Goldstein 6) L. Derobertis

ردهٔ توزیعهایی است که در مسأله به عنوان ردهٔ آلاینده در نظر گرفته می‌شود. گزینش Q، تأثیر بسیار مهمی دارد و غالباً با توجه به خصوصیات شکلی توزیع پیشین و چهار هدف عمده‌ای که هر رده باید داشته باشد، انتخاب می‌شود. رده‌های مختلفی که می‌توان به عنوان رده آلوده‌ساز در نظر گرفت، عبارت‌اند از $\theta_\circ$ (مد توزیع پیشین $\pi_\circ$):

$$Q_A = \{\text{تمام توزیعهایی مانند} q\} \quad (۱۰)$$

$$Q_{U\circ} = \quad (۱۱)$$

$$\{q : \text{تک مدی است } \pi = (1-\varepsilon)\pi_\circ + \varepsilon q \text{ پیشینی که به ازای آنها } q\}$$

$$Q_U = \{q : \theta_\circ \text{ متشکل از پیشینی تک مدی، با مد } q\} \quad (۱۲)$$

$$Q_{SU} = \{q : \theta_\circ \text{ تمام توزیعهای تک مدی و متقارن، با مد } q\} \quad (۱۳)$$

رده‌های (۱۰)، (۱۱)، (۱۲) برای حالتهایی طراحی شده‌اند که در آنها $\pi_\circ$ تک‌مدی است، و توزیع پیشین، به طورکلی تک مدی ارزیابی می‌شود. این گونه مشخصات شکلی به طورکلی و معمول، مناسب هستند.

به کارگیری ردهٔ Q_A بسیار آسان است، اما در این وضعیت Γ_ε شامل بسیاری توزیعهای نامعقول نیز هست (از جمله توزیعهای تباهیده که خیلی دور از $\pi_\circ$ هستند). بنابراین، $(\underline{\rho}(x,\pi),\bar{\rho}(x,\pi))$، در اکثر موارد بازهٔ بسیار بزرگی است. وقتی به تک مدی بودن پیشین معتقد هستیم، Γ_ε با $Q_{U\circ}$ بسیار معقول است. این رده شامل تمامی پیشینهای معقول است (که تک مدی و نزدیک به $\pi_\circ$ هستند) و هیچ توزیع پیشین غیرمعقول را شامل نمی‌شود. متأسفانه، محاسبهٔ بازهٔ تغییرات معیارهای پسین در این رده مشکل است

استفاده از Q_U یا Q_{SU} فاصلهٔ کوتاهتری نسبت به Q_A به دست می‌دهد، اگر چه امکان دارد که پسینهای معقول خاصی در نظر گرفته نشوند. توجه کنید، اگرچه Γ_ε شامل پیشینهای بسیاری است ولی به طور ویژه شامل توزیعهایی با دمهای کلفت‌تر از $\pi_\circ$ است. مزیت استفاده از Q_U یا Q_{US} نسبت به $Q_{U\circ}$، محاسبهٔ آسانتر معیارهای پسین است.

جهت آگاهی بیشتر با ردهٔ εـ آلوده‌ها، هوبر (۱۹۷۳)، برگرـ برلینر (۱۹۸۶)، برگر (۱۹۸۵)، سیواگنسن (۱۹۸۸)، برگر ـ سیواگنسن (۱۹۸۹) را مطالعه کنید.

۲.۳ روشهایی برای محاسبهٔ $\bar{\rho}(x,\pi)$ و $\rho(x,\pi)$

روشهای بیشینه و کمینه کردن $\rho(x,\pi)$ به ازای هر π در ردهٔ Γ معمولاً مختص همان معیار پسین در ردهٔ Γ هستند. در اکثر موارد ایدهٔ اساسی این روشها، تعیین زیر رده‌ای از Γ با بعد کوچکتر است، به طوری که پیشینهای بیشینه (کمینه) باید در این زیر ردهٔ $\Gamma_\circ$ واقع شوند. سپس بهینه کردن معیار مذکور فقط نیازمند محاسبات عددی در این زیر ردهٔ $\Gamma_\circ$ است (البته امکان دارد در مسأله خاصی روش معین دیگری مورد نیاز باشد). در این بخش چندین روش کلی را جهت بهینه کردن معیارهای ذکر شده در بخش ۱ـ۳ معرفی می‌کنیم.

۳.۳ تابعکهای خطی

کمینه و بیشینه کردن تابعکهای خطی از π به وضوح آسان است. در بسیاری از موارد Γ محدب است و یافتن نقاط اکسترمم در $\Gamma_\circ$ میسر است و

$$\sup_{\pi\in\Gamma} Q(\pi) = \sup_{s\in\Gamma}\int h(\theta)\pi(\theta)d\theta = \sup_{x\in\Gamma_\circ} Q(\pi)$$

$$\inf_{\pi\in\Gamma} Q(\pi) = \inf_{\pi\in\Gamma_\circ} Q(\pi). \quad (۱۴)$$

در روابط فوق $\Gamma_\circ$ مجموعه‌ای با بعد کوچکتر است، به طوری که با روشهای عددی به آسانی می‌توان حدود بالا و پایین معیار $Q(\pi)$ را به ازای هر π در ردهٔ Γ و متناظر با آن در زیر ردهٔ $\Gamma_\circ$ محاسبه نمود.

مثال ۱. ردهٔ Q_{SU} را در نظر بگیرید. فرض کنید معیار مورد علاقه، چگالی حاشیه‌ای $m(\pi)$ باشد. به وضوح داریم:

$$m(x|\pi) = \int_\Theta f(x|\theta)[(1-\varepsilon)\pi_\circ(\theta) + \varepsilon q(\theta)]d\theta$$
$$= (1-\varepsilon)m(x|\pi_\circ) + \varepsilon m(x|q). \quad (۱۵)$$

توجه کنید که می‌توان هر توزیع متقارن (حول $\theta_\circ$) و تک مدی مانند q را به صورت ترکیبی از توزیعهای یکنواخت متقارن نوشت. بنابراین نقاط اکسترمم Q_{SU} در واقع چگالیهای یکنواخت بر فاصلهٔ $(\theta_\circ - Z, \theta_\circ + Z)$ هستند، و

$$\Gamma_\circ = \{\text{نقاط اکسترمم} \Gamma\}$$
$$= \{\pi = (1-\varepsilon)\pi_\circ + \varepsilon U(\theta_\circ - Z, \theta_\circ + Z), Z > 0\}.$$

بنابراین

$$\sup_{\pi\in\Gamma} m(x,\pi) = (1-\varepsilon)m(x,\pi_\circ) + \varepsilon \sup_{Z>0}\int_{\theta_\circ - Z}^{\theta_\circ + Z} \frac{1}{2Z} f(x|\theta)d\theta)$$

در صورتی که ردهٔ آلوده‌سازها را ردهٔ Q_U در نظر بگیریم، خواهیم داشت:

$$\Gamma_\circ = \{\pi = (1-\varepsilon)\pi_\circ + \varepsilon U(\theta_\circ, \theta_\circ + Z), Z > 0\}$$

$$\sup_{\pi\in\Gamma} m(x|\pi) = (1-\varepsilon)m(x|\pi_\circ) + \varepsilon \sup_{Z>0}\int_{\theta_\circ}^{\theta_\circ + Z} \frac{1}{Z} f(x|\theta)d\theta,$$

و برای ردهٔ Q_A،

$$\sup_{\pi\in\Gamma} m(x|\pi) = m(x|\hat{\pi}),$$

که در آن $\hat{\pi}(0) = (1-\varepsilon)\pi_0 + \varepsilon\hat{q}$ و $\hat{q}$ پیشینی است که $m(x|q)$ را ماکسیمم می‌کند. $\hat{q}$ی پیشین $\hat{\pi}$ را پیشین $ML-II$ یا درستنمایی از نوع دوم می‌گوییم.

۴.۳ نسبت خطی معیارهای پسین

معیار پسین زیر را در نظر بگیرید:

$$\rho(x,\pi) = \frac{\int f(\theta)\pi(\theta)d\theta}{\int g(\theta)\pi(\theta)d\theta}. \tag{16}$$

که در آن $g(0) > 0$. در این صورت نتیجه می‌گیریم که

$$\sup_{\pi\in\Gamma}\rho(x,\pi) = \sup_{\theta\in\Theta}\frac{f(\theta)}{g(\theta)}, \inf_{\pi\in\Gamma}\rho(x,\pi) = \inf_{\theta\in\Theta}\frac{f(\theta)}{g(\theta)}. \tag{17}$$

این نتیجه‌ای استاندارد است که در برگر و سیواگنسن (۱۹۸۹) آمده است

برای ردهٔ بسیار رایج εـ آلوده‌ها نتیجهٔ جالب زیر به دست آمده است: وقتی‌که ردهٔ آلوده‌ها تمام توزیعهای ممکنه است ($Q_A = \{$تمام توزیعها$\}$) : برای هر $\pi \in \Gamma_\varepsilon$ با ردهٔ آلودگی Q_A، کمیت نسبت خطی معیارهای پسین را می‌توان به صورت زیر نوشت:

$$\begin{aligned}\rho(x,\pi) &= E^{\pi}[h(\theta)]\\ &= \frac{(1-\varepsilon)\int h(\theta)f(x|\theta)\pi_0(\theta)d\theta + \varepsilon\int h(\theta)f(x|\theta)q(\theta)}{(1-\varepsilon)\int f(x|\theta)\pi_0(\theta)d\theta + \varepsilon\int f(x|\theta)q(\theta)d\theta}\\ &= \frac{\int f(\theta)q(\theta)}{\int g(\theta)q(\theta)},\end{aligned}$$

که در آن،

$$f(\theta) = (1-\varepsilon)\int h(\theta)f(x|\xi)\pi_0(\xi)d\xi + \varepsilon h(\theta)f(x|\theta)$$

و

$$g(\theta) = (1-\varepsilon)\int f(x|\xi)\pi_0(\xi)d\xi + \varepsilon f(x,|\theta),$$

و با توجه به روابط (۱۵) و (۱۶)، می‌توان به سهولت $\underline{\rho}(x|\pi)$ و $\bar{\rho}(x|\pi)$ را مشخص نمود، و در واقع با بیشینه و کمینه کردن $\frac{f(\theta)}{g(\theta)}$ نسبت به θ، معیارهای مذکور از حل مسأله‌ای عددی به دست می‌آیند.

هوبر در سال ۱۹۷۳ این نتیجه را برای $h(\theta) = I_C(\theta)$، که در آن $\rho(x,\pi)$، احتمال پسین مجموعهٔ C گرفته شده است، بسط داد و سیواگنسن (۱۹۸۸) سایر معیارهای پسین را از قبیل میانگین و واریانس مورد بررسی قرار داد. در تمامی روشهای بالا، برای سادگی بیشتر در حل مسألهٔ عددی، حل مسأله به صورت حل معادلات تکراری به طور معمول مرسوم و میسر است.

۴ وقتی که ردهٔ آلاینده، توزیع‌های متقارن و تک مدی است ($Q = Q_{SU}$)

اغلب اوقات باورهای پیشین به گونه‌ای است که معقول به نظر می‌رسد Q را در ردهٔ Γ_ε، ردهٔ توزیعهای تک مدی و متقارن حول θ_0 در نظر بگیریم (به عبارت دیگر $Q = Q_{SU}$). وقتی که π نیز متقارن است، این تصمیم معقولتر به نظر می‌آید. با در نظرگرفتن این رده به عنوان ردهٔ آلاینده می‌توانیم به بررسی نیرومندی «مینیمال» بپردازیم. به عبارت دیگر با انتخاب این رده به عنوان ردهٔ آلاینده، می‌توان حساسیت صورت تابعی دقیقی از پیشین و دنباله‌های پیشین را مورد تحقیق قرار داد. (در مقابل، انتخاب ردهٔ Q_A منجر به بررسی نیرومندی «ماکسیمال» می‌گردد؛ برگر و سیواگنسن (۱۹۸۹) این مسأله را حل کردند.) هر $q \in Q_{SU}$ را می‌توان به صورت ترکیبی از یکنواخت نوشت، یعنی

$$q(\theta) = \int_0^\infty \frac{1}{2z} I_{(\theta_0 - z, \theta_0 + z)}(\theta) G(z) dz,$$

که G یک توزیع دلخواه بر $(0,\infty)$ است. هر نسبت خطی از معیارهای پسین را می‌توان به صورت زیر نوشت:

$$\begin{aligned}\rho(x,\pi) &= E^{\pi}[h(\theta)]\\ &= \frac{(1-\varepsilon)\int h(\theta)f(x|\theta)\pi_0(\theta)d\theta + \varepsilon\int_0^\infty H_1(z)G(z)dz}{(1-\varepsilon)\int f(x|\theta)\pi_0(\theta)d\theta + \varepsilon\int_0^\infty H_2(z)G(z)dz}\end{aligned}$$

که در آن

$$\begin{aligned}H_1(z) &= \frac{1}{2z}\int_{\theta_0 - z}^{\theta_0 + z} h(\theta)f(x|\theta)d\theta)d\theta,\\ H_2(z) &= \frac{1}{2z}\int_{\theta_0 - z}^{\theta_0 + z} f(x|\theta)d\theta.\end{aligned}$$

اکنون تعریف می‌کنیم:

$$\begin{aligned}f(z) &= (1-\varepsilon)\int h(\xi)f(x|\xi)\pi_0(\xi)d\xi + \varepsilon_1 H_1(z)\\ g(z) &= (1-\varepsilon)\int f(x|\xi)\pi_0(\xi)d\xi + \varepsilon_2 H_2(z).\end{aligned}$$

بنابراین

$$\rho(x,\pi) = \frac{\int f(z)G(z)}{\int g(z)G(z)dz},$$

و با توجه به روابط (۱۵) و (۱۶)، $\underline{\rho}(x,\pi)$ و $\bar{\rho}(x,\pi)$ متناظر با ماکسیمم و مینیمم تابع $\frac{f(z)}{g(z)}$ نسبت به z هستند.

تحلیل برای ردهٔ $Q = Q_U$ شبیه به ردهٔ Q_{SU} است، نتایج این بررسی

در برگر و سیواگنسن (۱۹۸۹) آمده است. تحلیل برای ردهٔ Q_{U° بسیار مشکل است. چرا که Q_{U° به گونه‌ای تعریف شده است که قیدی بسیار کلی بر روی پیشین قرار داده شده است. تعیین $\bar{\rho}(x,\pi)$ و $\underline{\rho}(x,\pi)$ تحت Q_{U°، وقتی که ρ چگالی حاشیه‌ای در برگر و برلینر (۱۹۸۶)، یا میانگین پسین باشد در سیواگنسن (۱۹۸۹) آمده است. ارائهٔ جزئیات در این مقاله بسیار پیچیده است، اگر چه روش اساسی ماکسیمم کردن برای $\pi \in \Gamma$ به طور خلاصه عبارت است از:

π در داخل بازه، تابعی پیوسته و یکنواخت است و در خارج از این بازه برابر با $(1-\varepsilon)\pi_\circ(\theta)$ است.

۱.۴ نکاتی چند در مورد گزینش Q

همان گونه که قبلاً نشان دادیم، انتخاب Γ (در واقع) مستلزم برخورداری از چندین ویژگی است. به کار بردن Q_A محاسباتی ساده را به همراه دارد، و اطمینان داریم که هیچ پیشین معقولی نادیده گرفته نمی‌شود.

خصوصیات منفی ردهٔ Q_A این است که شامل بسیاری از توزیعهایی است که بدون شک غیرمعقول هستند، و باعث می‌شود بازهٔ $\bar{\rho}(x,\pi)$ و $\underline{\rho}(x,\pi)$ بسیار وسیعتر از حد معمول شود. از طرف دیگر وقتی از ردهٔ Q_A استفاده می‌کنیم، در مورد مسائل حدی، نتیجه‌ای نادرست به دست می‌آید (سیواگنسن (۱۹۸۸) و برگر ـ برلینر (۱۹۸۶) را مطالعه کنید).

با توجه به موارد فوق‌الذکر، تحدید Q بسیار معقول به نظر می‌رسد. رده‌های Q_{U°، Q_U، و Q_{SU} همگی مناسب هستند و می‌توان نشان داد که به طور مجانبی نتیجه‌ای مناسب و مقتضی حاصل می‌کنند. مثال زیر، مقایسه‌ای بین بازه‌های $\bar{\rho}(x,\pi)$ و $\underline{\rho}(x,\pi)$ برای رده‌های مختلف را نشان می‌دهد.

مثال ۲) فرض کنید که $X \sim N(\theta, 1)$ و $\pi_\circ$ توزیع نرمال استاندارد باشد، $\varepsilon = 0/1$. برای هر یک از چهار ردهٔ مذکور برای مقادیر مختلف x، دامنه $(\underline{\rho}, \bar{\rho})$ از میانگین پسین برای θ در جدول زیر بیان می‌شود. دقّت شود که با توجه به توضیح فوق، حدود دامنهٔ برای Q_{SU} اساساً کوچکتر از حدود دامنه Q_A هستند. همچنین توجه کنید x بزرگتر، متناظر با ناسازگاری بین داده‌ها و $\pi_\circ$ است، و بنابراین عدم حتمیت بیشتر را نتیجه می‌گیریم. از نقطه نظر محاسباتی، تحلیل با Q_{SU} و Q_n هر دو بسیار آسان هستند، Q_{U° اساساً بسیار مشکل است. همچنین انتخاب $\pi_\circ$، به ویژه در رفتار دمهایش، می‌تواند تأثیر مشخص و به سزایی داشته باشد. برای آگاهی بیشتر در این زمینه برگر (۱۹۹۰) برگر ـ برلینر (۱۹۸۶)، برگر ـ سیواگنسن (۱۹۸۹) را مطالعه کنید.

۵ نتیجه‌گیری

نیرومندی بیزی یکی از روشهایی است که برای پاسخ دادن به شکل تعیین توزیع پیشین مطرح شده است. مسلماً در کلیهٔ مسائل بیزی که دفاع از یک پیشین به کار رفته، مشکل است یا دانشی ناقص برای تعیین پیشین موجود است، قابل ارائه می‌باشد. در این مقاله، مسأله و روشهای عام برخورد با آن را مطرح کردیم. کاربرد آن در مباحث رگرسیون، طرحهای آزمایشی، سریهای زمانی، و حوزه‌های متعدد دیگر می‌تواند به نتایج با ارزشی منجر شود.

مراجع

[1] Berger, J. (1984). The robust Bayesian viewpoint (with discussion). In: J. Kadane, Ed., Robustness of Bayesian Analysis. North-Holland, Amsterdam.

[2] Berger, J. (1985). Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York.

[3] Berger, J. and L. M. Berliner (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. Ann. statist. 14, 461-186.

[4] Deroberts, L. and Hartigan, J. A. (1981). Bayesian Inference Using Intervals of Measures. The Annals of Stutistics, Vol. 9, No. 2, 235-244.

[5] Berger, J. (1990) Robust Bayesian analysis; Sensitivity to the prior. J. Statis. Palnning and inf. 25, 303-328.

[6] Goldstein, M. (1980). The linear Bayes regression estimator under weak prior assumptions, *Biometrica*, 67, 621-628.

[7] Good, I. J. (1952). Rational decisions. J. Roy. Statist. Soc. Ser. B 14, 107-114.

[8] Hartigar, J. A. (1969). Linear Bayesian Methods, J. Roy. Statist. Soc. Ser. B31, 446-454.

[9] Huber, P. J. (1973). The use of Choquet capacities in statistics. Bull. Inst. Internal. Statist. 45, 181-191.

[10] Sivaganesan, S. (1988). Ranges of posterior measures for priors with arbitrary contaminations. Comm. Statist. 17, 1591-1612.

[11] Sivaganesan, S. and J. Berger (1989). Ranges of posterior measures for priors with unimodal contaminations. Ann. Statist. 17, 868-889.

[12] Stone, M. (1963). Robustness of nonideal decision procedures. J. Amer. Statist. Assoc. 58, 480-486.

[۱۳] محمودی بایبوری، همایون (۱۳۷۰). «نیرومندی بیزی و کاربرد آن در آزمون همگنی برای جدولهای دوجمله‌ای» پایان‌نامه کارشناسی‌ارشد، دانشگاه شهید بهشتی.

[۱۴] دانشخواه، علیرضا (۱۳۷۵). «نیرومندی بیزی در نمونه‌گیری از جامعه‌های متناهی» پایان‌نامهٔ کارشناسی‌ارشد، دانشگاه شهید بهشتی.

نخستین کلاسهای آماری

«صبح امروز، کنفرانسی در دفتر اقای سعید سمیعی رئیس کل آمار عمومی با شرکت روسای آمار وزارت خانه‌ها تشکیل گردید. دراین جلسه موضوع تعلیم و تربیت آمارشناسان و آمارگران مورد مذاکره قرار گرفت و رئیس کل آمار عمومی به آقایان روسای ادارات آمار وزارتخانه‌ها خاطرنشان کرد که چون در مهرماه سال جاری تدریس علم آمار در دانشگاه تهران شروع می‌شود، از هم‌اکنون روسای آمار وزارتخانه‌ها باید به وسیله تعلیمات نظری و علمی در طی کنفرانس‌ها و اجرای بازدید از دستگاههای استخراج آمار که برنامه آن در همین دو روزه از طرف آمار عمومی در اختیار آنان گذاشته خواهد شد، عده‌ای از کارمندان ادارات آمار وزارتخانه‌ها را که مستعد شرکت در کلاس آمار هستند مجهز و آماده نمایند.»

این خبر، مندرج در روزنامهٔ اطلاعات روز یکشنبه اول تیرماه سال ۱۳۳۷ و به نقل از روزنامهٔ اطلاعات روز یکشنبه ۳۱ خرداد ۱۳۷۷، حاکی از دایر شدن نخستین کلاسهای درس آمار در ایران در ۴۰ سال قبل است. جالب است بدانیم که نخستین کلاسهای درس آمار در جایی که اینک بهترین گروههای آماری دنیا را دارد، یعنی کشور ایالات متحده آمریکا، در سال ۱۹۱۸ میلادی، یعنی ۸۰ سال قبل، دایر شده‌اند.

مطالب زیر به نقل از مقالهٔ «تاریخچهٔ پیدایش آمار نوین در آمریکا» تألیف بوید هارشبارگر (Boyd Harshbarger) مندرج در "On the History of Statistics and Probability" به ویراستاری اوون (D. B. Owen) شاهد این مدعاست:

«موضوع آمار به گونه‌ای که امروزه در ذهن ماست، در ایالات متحده ریشه در عصر حاضر دارد. پروفسور الن کرگ (Allen T. Craig) در مقاله‌اش با عنوان «جشن بیست و پنجمین سالگی» که در Annals of Mathematical Statistics (۱۹۶۰) منتشر شد، اظهار می‌دارد که «پیش از سال ۱۹۲۰ تعداد انگشت شماری از مدارس عالی و دانشگاههای آمریکا، هرکس را که علاقه‌ای جدی به روشهای استنباط علمی جدیداً در حال ظهور با عنوان آمار ریاضی نشان می‌داد، به عضویت گروهها در می‌آوردند ... انجمن ریاضی آمریکایی عالی جاه نسبت به کل این ماجرا نظر بدبینانه‌ای داشت و به این افراد تکرو (آماردانان) با تردیدی توأم با سوءظن می‌نگریست.»

پروفسور ریتس (H. L. Rietz)، در دانشگاه آیووا (۱۹۱۸)، درسهایی در آمار دایر کرد که آمار ریاضی هم در بین آنها بود. به علاوه وی یک تک نگاری آمار ریاضی (۱۹۲۷) را به کشور تقدیم کرد که پیشگام آمار نوین بود. میزان موفقیت وی در این زمینه را می‌توان با ملاحظه نام دانشجویانی که زیرنظر او درس خواندند و آنها که مدرک خود را در این رشته از دانشگاه آیووا دریافت کردند، سنجید. می‌توان از بین آنها به پروفسور فرانک وایدا (دکترا در سال ۱۹۲۳) از دانشگاه پرینستون، پروفسور آلن کرگ (دکترا در سال ۱۹۳۱) از دانشگاه آیووا، پروفسور جان کرتیس (دکترا در سال ۱۹۳۰) از دانشگاه کورنل، و پروفسور هربرت مه‌یرز از دانشگاه فلوریدا، اشاره کرد. ...»

م. ق. و.