آنتروپی و کاربرد آن در تشخیص

ناهید سنجری[۱] مهدی زارعی[۲]

چکیده

هدف اصلی این مقاله معرفی و تحلیل برخی از خواص تابع کالبک-لیبلر برای انجام تشخیص بین توزیعها می‌باشد. برای این منظور شاخصی جهت اندازه‌گیری میزان اختلاف بین توزیعها با توجه به تابع کالبک-لیبلر و اصل ماکزیمم آنتروپی معرفی می‌شود.

۱ مقدمه

آنتروپی متغیر تصادفی X با تابع چگالی $f(X)$ به صورت

$$H(X) = -\int_{-\infty}^{\infty} f(x) Ln f(x) dx$$

تعریف می‌شود که از آن تحت عنوان میانگین عدم قطعیت مرتبط با متغیر تصادفی X یاد می‌شود[۴]. در این رابطه یکی از توابع مهم که کاربرد فراوانی در انجام مقایسه بین توزیعها دارد تابع اطلاع کالبک-لیبلر است. این تابع عبارتست از آنتروپی چگالی f_1 نسبت به چگالی f_2، که به نام آنتروپی نسبی هم مشهور است و به صورت

$$K(f_1 : f_2) = \int_{-\infty}^{\infty} f_1(x) Ln \frac{f_1(x)}{f_2(x)} dx$$

تعریف می‌شود. در بخش دوم مفهوم آنتروپی و اصل ماکزیمم آنتروپی بررسی می‌شود. در بخش سوم تابع کالبک-لیبلر به صورت کامل معرفی شده و شاخصی تحت عنوان شاخص اطلاع ID (Information Discrimination) برای اندازه‌گیری میزان اختلاف بین توزیعها مطرح شده است. در بخش چهارم آماره‌های (ID) معرفی می‌شوند.

۲ آنتروپی و اصل ماکزیمم آنتروپی

آنتروپی در حالت گسسته بیانگر عدم قطعیت مورد انتظار مرتبط با برآمد یک مشاهده از متغیر تصادفی گسسته X است. به عبارت دیگر هرگاه $P(x) = P(X = x)$ تابع جرم احتمال متغیر تصادفی X باشد آنتروپی متغیر تصادفی X عبارتست از

$$H(X) = -\sum_{x} P(x) Log P(x)$$

[۱] ناهید سنجری فارسی‌پور، بخش آمار، دانشگاه شیراز
[۲] مهدی زارعی، بخش آمار، دانشگاه شیراز

می‌توان نشان داد که $H(X)$ در این حالت همواره عددی مثبت است که بنابر پایه Log با واحدهای مختلفی سنجیده می‌شود. آنتروپی تابعی از احتمالات بوده و به آسانی مشاهده می‌شود که وقتی همه احتمالات برابرند (حالت یکنواخت) $H(X)$ بیشترین مقدار را خواهد داشت که به مفهوم کمترین اطلاع و بیشترین عدم قطعیت در مورد برآمد X است و در مقابل اگر شاهد احتمال ۱ $= P(X=x)$ باشیم $H(X)$ صفر است. یعنی عدم قطعیتی در مورد برآمد X وجود ندارد. البته تفسیر آنتروپی و مقدار بدست آمده برای آن در حالت گسسته در این نوشتار مورد نظر نمی‌باشد بلکه توجه اصلی به تعریف آنتروپی در حالتی است که X یک متغیر تصادفی پیوسته با تابع چگالی $f(x)$ می‌باشد که از آن به عنوان differential entropy یاد شده و به صورت زیر تعریف می‌شود

$$H(X) = -\int_{-\infty}^{\infty} f(x) Lnf(x)dx = -E[Lnf(X)] \quad (۱)$$

باید توجه داشت که در این حالت لزومی ندارد $H(X)$ مثبت باشد.

مثال: فرض کنید متغیر تصادفی X دارای توزیع نرمال با میانگین صفر و واریانس $\sigma^۲$ و تابع چگالی

$$f(x) = (۲\pi\sigma^۲)^{-\frac{۱}{۲}} \exp(-x^۲/۲\sigma^۲)\ ,\ -\infty < x < \infty$$

باشد. بنابر (۱)

$$\begin{aligned} H(X) &= -\int_{-\infty}^{\infty} f(x) Lnf(x)dx \\ &= -\int_{-\infty}^{\infty} f(x)[-Ln(۲\pi\sigma^۲)^{\frac{۱}{۲}} - x^۲/۲\sigma^۲]dx \\ &= \frac{۱}{۲} Ln ۲\pi e\sigma^۲ \end{aligned}$$

یک استنباط سریع از نتیجه بدست آمده حاکی از آن است که با زیاد شدن پراکندگی، آنتروپی افزایش می‌یابد.

برای اکثر توزیعهای پیوسته، آنتروپی محاسبه شده و در جداولی تنظیم گردیده است[۶].

اصل ماکزیمم آنتروپی، خانواده‌ای از توزیعها را به صورت

$$\Omega_\theta = \{f(x\mid\theta) : E_f[T_j(X)] = \theta_j,\ j = ۱,\ldots,m\}$$

در نظر می‌گیرد که T_jها توابعی هستند که با توجه به f انتگرال‌پذیر می‌باشند[۳].

برای حالت پیوسته، استنباط براساس مدلی که با توجه به قیود تعریف شده در Ω_θ آنتروپی (۱) را ماکزیمم سازد بنا نهاده شده است.

مدل ماکزیمم آنتروپی $f^*(x\mid\theta)$ در Ω_θ، اگر وجود داشته باشد، به شکل کلی

$$f^*(x\mid\theta) = M(\theta)\exp[\lambda_۱(\theta)T_۱(x) + \ldots + \lambda_m(\theta)T_m(x)]$$

می‌باشد که در آن $M(\theta)$ ثابت نرمال ساز بوده و $\lambda_۱,\ldots,\lambda_m$ به گونه‌ای انتخاب می‌شوند که $f^*(x\mid\theta)$ قیود موجود در Ω_θ را برآورده سازد. این شکل کلی را می‌توان با ظاهری بسته‌تر به صورت

$$f^*(x\mid\theta) = \exp\{\lambda_\circ + \lambda_۱T_۱(x) + \ldots + \lambda_mT_m(x)\} \quad (۲)$$

نشان داد. بسیاری از توزیعهای معروف، با توجه به قیود گوناگون ME (ماکزیمم آنتروپی) هستند که به همین منظور جداولی نیز تهیه شده است[۵].

مثال: اگر قید $\int_\circ^\infty xf(x)dx = \mu$ داده شده باشد در این صورت $T_۱(X) = X$ و $\theta_۱ = \mu$. بنابراین مسئله یافتن توزیع ME در خانواده

$$\Omega_\mu = \{f(x\mid\mu) : E(X\mid\mu) = \mu\}$$

به صورت پیدا کردن چگالی احتمالی است که با توجه به قید $\int_\circ^\infty xf(x)dx = \mu$ دارای بزرگترین آنتروپی باشد. با توجه به شکل کلی (۲) یعنی

$$۲f^*(x) = e^{\lambda_\circ + \lambda_۱ x}$$

مشاهده می‌شود که تنها تابع چگالی که قید بالا را برقرار می‌سازد چگالی نمایی است و بنابراین

$$f^*(x) = \frac{۱}{\mu}e^{-x/\mu}\ ,\ x \geq \circ$$

۳ شاخص اطلاع ID

همانگونه که اشاره شد آنتروپی چگالی f_1 نسبت به چگال f_2 (آنتروپی نسبی) تحت عنوان تابع کالبک–لیبلر به صورت

$$K(f_1 : f_2) = \int_{-\infty}^{\infty} f_1(x) Ln \frac{f_1(x)}{f_2(x)} dx \qquad (۳)$$

تعریف می‌شود و فرض می‌کنیم چگالی‌های f_1 و f_2 به گونه‌ای باشند که انتگرال وجود داشته باشد. در حقیقت تابع کالبک–لیبلر اندازه‌ای از فاصله یا اختلاف بین دو توزیع می‌باشد[۳].

یکی از خواص مهم تابع کالبک–لیبلر که برای دو حالت گسسته و پیوسته یکسان بوده و به آسانی قابل اثبات است [۱]، بیان می‌کند که اگر f_1 و f_2 دو تابع چگالی باشند آنگاه

$$K(f_1 : f_2) \geq 0$$

و تساوی تنها وقتی برقرار است که $f_1(x) = f_2(x)$ برای تمام xها.

بنابراین با بزرگ شدن مقدار $K(f_1 : f_2)$ اختلاف کمتری را انتظار داریم. با توجه به این مطلب توزیعها در خانواده

$$\Omega_\theta = \{f(x \mid \theta) : E_f[T_j(X)] = \theta_j \ , \ j = 1, \ldots, m\}$$

را می‌توان با توجه به تابع کالبک–لیبلر و توزیع ME به صورت

$$K(f : f^*) = \int_{-\infty}^{\infty} f(x \mid \theta) Ln \frac{f(x \mid \theta)}{f^*(x \mid \theta)} dx \qquad (۴)$$

مقایسه نمود. حالت $K(f : f^*) = 0$ وقتی برقرار است که برابری $f(x \mid \theta) = f^*(x \mid \theta)$ برای تمام xها برقرار باشد. همچنین با بسط عبارت (۴) می‌توان نوشت:

$$K(f : f^*) = \int_{-\infty}^{\infty} f(x \mid \theta) Lnf(x \mid \theta) dx - \int_{-\infty}^{\infty} f(x \mid \theta) Lnf^*(x \mid \theta) dx$$

با توجه به آنتروپی (۱) و شکل کلی $f^*(x \mid \theta)$ داریم:

$$\begin{aligned} K(f : f^*) &= -H[f(x \mid \theta)] \\ &\quad -\lambda_0 - \lambda_1 E_f[T_1(X)] - \ldots - \lambda_m E_f[T_m(X)] \end{aligned}$$

از طرفی با توجه به این که $E_f[T_j(X)] = \theta_j$ و نیز آنتروپی توزیع ME یعنی

$$H[f^*(x \mid \theta)] = -\lambda_0 - \lambda_1\theta_1 - \ldots - \lambda_m\theta_m$$

تابع کالبک–لیبلر را می‌توان به صورت

$$K(f : f^*) = H[f^*(x \mid \theta)] - H[f(x \mid \theta)]$$

نوشت. در واقع موضوع مورد علاقه، چگونگی دقت تقریب $f^*(x \mid \theta)$ از توزیع مولد داده‌ها می‌باشد. در صورتی که توزیع مولد داده‌ها بتواند به طور رضایتبخشی با استفاده از قیود اطلاع تعیین شده برای اعضای Ω_θ توصیف شود، انتظار می‌رود که آنتروپی آن تا اندازه‌ای به ME داده شده به صورت $H[f^*(x \mid \theta)]$ نزدیک باشد. بنابراین می‌توان شاخص ID را برای انجام مقایسه بین توزیعها براساس تابع کالبک–لیبلر به صورت

$$\begin{aligned} ID(f : f^* \mid \theta) &= 1 - \exp[-K(f : f^* \mid \theta)] \\ &= 1 - \exp\{H[f(x \mid \theta)] - H[f^*(x \mid \theta)]\} \end{aligned}$$

تعریف نمود. دو توزیع $f_1(x \mid \theta)$ و $f_2(x \mid \theta)$ در Ω_θ را ID تشخیص‌پذیر گوئیم اگر

$$ID(f_1 : f^* \mid \theta) \neq ID(f_2 : f^* \mid \theta)$$

باشد. با توجه به این نکته که $K(f : f^*) \geq 0$ است مشاهده می‌شود:

$$0 \leq ID(f : f^* \mid \theta) \leq 1$$

و یک مقدار $ID(f : f^* \mid \theta)$ نزدیک صفر بیانگر آن است که f و f^* تقریباً مشابه‌اند. این بدان معنی است که قیود استفاده شده برای محاسبه ME مقدار اطلاع بالایی برای توزیع f دربر دارند[۵]. در مثال بعد ID تشخیص‌پذیری چند خانواده مقیاس–مکان را نسبت به نرمال بودن بررسی می‌کنیم.

مثال: اگر $T_1(X) = X^۲$ و $\theta_1 = \sigma^۲$ باشد آنگاه توزیع ME در خانواده

$$\Omega_\sigma = \{f(x \mid \sigma^۲) : E(X^۲) = \sigma^۲\}$$

با توجه به (۲) توزیع نرمال $f^*(x \mid \sigma^۲) = N(۰, \sigma^۲)$ می‌باشد. خانواده‌های مقیاس—مکان زیر را در نظر بگیرید:

الف) لجستیک با چگالی

$$\begin{aligned} f_1(x \mid \theta_1, \lambda_1) &= \lambda_1 \exp\{-\lambda_1(x-\theta_1)\} \\ &\quad \times[۱ + \exp\{-\lambda_1(x-\theta_1)\}]^۲, \\ &\quad -\infty < \theta_1 < \infty \ , \ \lambda_1 > ۰ \end{aligned}$$

ب) لاپلاس با چگالی

$$\begin{aligned} f_۲(x \mid \theta_۲, \lambda_۲) = \frac{۱}{۲}\lambda_۲ \exp\{-\lambda_۲ \mid x-\theta_۲ \mid\}, \\ -\infty < \theta_۲ < \infty, \lambda_۲ > ۰ \end{aligned}$$

ج) گمبل با چگالی

$$\begin{aligned} f_۳(x \mid \theta_۳, \lambda_۳) &= \lambda_۳ \exp\{-\lambda_۳(x-\theta_۳)\} \\ &\quad \times \exp[-\exp\{-\lambda_۳(x-\theta_۳)\}], \\ &\quad -\infty < \theta_۳ < \infty \ , \ \lambda_۳ > ۰ \end{aligned}$$

θ_i و λ_i، $i = ۱,۲,۳$ پارامترهای مکان و مقیاس توزیعها می‌باشند. برای

$$\pi^۲/(۳\lambda_1^۲) = ۲/\lambda_۲^۲ = \pi^۲/(۲\lambda_۳^۲) = \sigma^۲$$

f_1 و $f_۲$ و $f_۳$ اعضای Ω_σ می‌باشند. آنتروپی مقیاس مکان بالا با واریانس متناهی $\sigma^۲$ به صورت

$$\begin{aligned} H[f_1(x \mid \sigma^۲)] &= \frac{۱}{۲}Ln(\sigma^۲) + \frac{۱}{۲}Ln(۳e^۴/\pi^۲) \\ H[f_۲(x \mid \sigma^۲)] &= \frac{۱}{۲}Ln(\sigma^۲) + \frac{۱}{۲}Ln(۲e^۲) \\ H[f_۳(x \mid \sigma^۲)] &= \frac{۱}{۲}Ln(\sigma^۲) + \frac{۱}{۲}Ln(۶e^۲/\pi^۲) + \gamma \end{aligned}$$

می‌باشد که $\gamma = ۰/۵۷۷۱$ ثابت اولر (Euler Constant) می‌باشد.

با به کارگیری ماکزیمم آنتروپی یعنی

$$H[f^*(x \mid \sigma^۲)] = \frac{۱}{۲}Ln(\sigma^۲) + \frac{۱}{۲}Ln(۲\pi e)$$

در عبارت

$$ID(f : f^* \mid \theta) = ۱ - \exp\{H[f(x \mid \theta)] - H[f^*(x \mid \theta)]\}$$

شاخصهای ID خانواده‌های مقیاس—مکان بالا نسبت به نرمال بودن به صورت زیر بدست می‌آیند:

$$\begin{aligned} ID(f_1 : f^* \mid \theta) &\approx ۰/۰۱۴ \\ ID(f_۲ : f^* \mid \theta) &\approx ۰/۰۷۰ \\ ID(f_۳ : f^* \mid \theta) &\approx ۰/۰۸۷ \end{aligned}$$

مشاهده شد که هر چقدر $ID(f : f^*)$ به صفر نزدیکتر باشد f و f^* شباهت بیشتری با یکدیگر دارند. بنابراین ملاحظه می‌شود که برای یک واریانس مشخص، توزیع لجستیک نزدیکترین خانواده به نرمال در بین سه خانواده مقیاس—مکان و گمبل دورترین آنها می‌باشد.

۴ آماره‌های ID

در بخش قبل شاخص ID را براساس تابع کالبک—لیبلر به صورت

$$ID(f : f^* \mid \theta) = ۱ - \exp\{H[f(x \mid \theta)] - H[f^*(x \mid \theta)]\} \quad (۵)$$

معرفی نمودیم. اکنون سعی می‌کنیم برآوردی از شاخص ID را ارائه دهیم. با توجه به عبارت (۵) برآورد $ID(f : f^* \mid \theta)$ موکول به برآورد $H[f(x \mid \theta)]$ و $H[f^*(x \mid \theta)]$ می‌باشد. ماکزیمم آنتروپی به صورت:

$$H[f^*(x \mid \theta)] = -LnM(\theta) - \lambda_1(\theta)\theta_1 - \ldots - \lambda_m(\theta)\theta_m$$

تابعی از θ است. وقتی که پارامتر θ مقدار مشخص $\theta_۰$ باشد، $H[f^*(x \mid \theta)]$ مقدار معلومی خواهد بود. ولی در بیشتر مسائل، θ براساس داده‌ها برآورد شده و بنابراین برآورد $H[f^*(x \mid \theta)]$ به یک مسئله برآورد پارامتر تبدیل می‌شود. در صورت داشتن

یک نمونه به حجم n، عبارت $H[f^*(x \mid \theta)]$ را می‌توان به طور مثال با قرار دادن برآورد حداکثر درستنمایی θ(MLE) تحت مدل ME، یعنی θ_n، برآورد نمود؛ در این صورت برآورد حداکثر درستنمایی آن یعنی $H[f^*(x \mid \theta)_n]$ به دست می‌آید. راهکار مدل سازی ME مستلزم برآورد آنتروپی توزیع مولد داده‌ها با بکارگیری یک روش ناپارامتری و مقایسه آن با آنتروپی $f^*(x \mid \theta)$ می‌باشد. رویه‌های ناپارامتری گوناگونی برای برآورد آنتروپی موجود است که از جمله می‌توان به برآورد ابراهیمی و دیگران (۱۹۹۴) اشاره کرد[۲].

استفاده از یک برآورد آنتروپی $H[f^*(x \mid \theta)_n]$ و یک برآورد ناپارامتری آنتروپی $H_n[f]$ در (۵) یک آماره ID را به صورت زیر بدست می‌دهد

$$ID_n(f : f^* \mid \theta_n) = ۱ - \exp\{H_n[f] - H[f^*(x \mid \theta_n)]\}$$

که $۱ \leq ID_n(f : f^* \mid \theta_n) \leq ۰$ بوده و در صورتی که مقدار $ID_n(f : f^* \mid \theta_n)$ نزدیک صفر نباشد، باید قیود اضافی یا انواع دیگری از قیود را جستجو نمود[۵].

مراجع

[1] Cover,T.M. and J.A.Thomas (1991) Elements of Information Theoty. New York: wiley.

[2] Ebrahimi,N. et al (1994) Two Measures of sample Entropy. statistics and probability letters, 20,225-234.

[3] Kullback,S. (1959) Information theory and Statistics. New York: wiley.

[4] Soofi,E.S (1994) Capturing the Intangible Concept of Information. JASA, 89, 1243-1254.

[5] Soofi,E.S. et al (1995) Information Distinguishability with Application to Analysis of Failure Data. JASA, 90, 657-668.

[6] Verdugo Lazo,A.C.G. and P.N.Rathie (1987) on the Entropy of Continous Probability Distributions. IEEE. trans on Info. Theory IT- 24, 120-122.