

آنتروپی و کاربرد آن در تشخیص

ناهید سنجرى^۱ مهدى زارعى^۲

چکیده

هدف اصلی این مقاله معرفی و تحلیل برخی از خواص تابع کالبک-لیبلر برای انجام تشخیص بین توزیعها می باشد. برای این منظور شاخصی جهت اندازه گیری میزان اختلاف بین توزیعها با توجه به تابع کالبک-لیبلر و اصل ماکزیمم آنتروپی معرفی می شود.

۱ مقدمه

صورت کامل معرفی شده و شاخصی تحت عنوان شاخص اطلاع ID (Information Discrimination) برای اندازه گیری میزان اختلاف بین توزیعها مطرح شده است. در بخش چهارم آماره های (ID) معرفی می شوند.

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

$$H(X) = - \sum_x P(x) \log P(x)$$

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

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

تعریف می شود که از آن تحت عنوان میانگین عدم قطعیت مرتبط با متغیر تصادفی X یاد می شود [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 می شود. در بخش دوم مفهوم آنتروپی و اصل ماکزیمم آنتروپی بررسی می شود. در بخش سوم تابع کالبک-لیبلر به

^۱ ناهید سنجرى فارسى پور، بخش آمار، دانشگاه شیراز
^۲ مهدى زارعى، بخش آمار، دانشگاه شیراز

در نظر می‌گیرد که T_j ها توابعی هستند که با توجه به f انتگرال پذیر 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 $f^*(x|\theta)$ در Ω_θ ، اگر وجود داشته باشد، به شکل کلی

$$f^*(x|\theta) = M(\theta) \exp[\lambda_1(\theta)T_1(x) + \dots + \lambda_m(\theta)T_m(x)]$$

می‌باشد که در آن $M(\theta)$ ثابت نرمال ساز بوده و $\lambda_1, \dots, \lambda_m$ به گونه‌ای انتخاب می‌شوند که $f^*(x|\theta)$ قیود موجود در Ω_θ را برآورده سازد. این شکل کلی را می‌توان با ظاهری بسته‌تر به صورت

$$f^*(x|\theta) = \exp\{\lambda_0 + \lambda_1 T_1(x) + \dots + \lambda_m T_m(x)\} \quad (2)$$

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 $\int_0^\infty x f(x) dx = \mu$ داده شده باشد در این صورت $T_1(X) = X$ و $\theta_1 = \mu$. بنابراین مسئله یافتن توزیع ME در خانواده

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

به صورت پیدا کردن چگالی احتمالی است که با توجه به قید $\int_0^\infty x f(x) dx = \mu$ دارای بزرگترین آنتروپی باشد. با توجه به شکل کلی (۲) یعنی

$$f^*(x) = e^{\lambda_0 + \lambda_1 x}$$

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

$$f^*(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0$$

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 $1 = P(X=x)$ 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

$$H(X) = - \int_{-\infty}^{\infty} f(x) \ln f(x) dx = -E[\ln f(X)] \quad (1)$$

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 σ^2 و تابع چگالی

$$f(x) = (\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2), \quad -\infty < x < \infty$$

باشد. بنابر (۱)

$$\begin{aligned} H(X) &= - \int_{-\infty}^{\infty} f(x) \ln f(x) dx \\ &= - \int_{-\infty}^{\infty} f(x) [-\ln(\pi\sigma^2)^{1/2} - x^2/2\sigma^2] dx \\ &= \frac{1}{2} \ln \pi e \sigma^2 \end{aligned}$$

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

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

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

$$K(f_1 : f_2) = \int_{-\infty}^{\infty} f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx \quad (3)$$

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 x ها.

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

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

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

$$K(f : f^*) = \int_{-\infty}^{\infty} f(x|\theta) \ln \frac{f(x|\theta)}{f^*(x|\theta)} dx \quad (4)$$

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

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

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

$$K(f : f^*) = -H[f(x|\theta)] - \lambda_0 - \lambda_1 E_f[T_1(X)] - \dots - \lambda_m E_f[T_m(X)]$$

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

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

تابع کالبک-لیبلر را می‌توان به صورت

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

نوشت. در واقع موضوع مورد علاقه، چگونگی دقت تقریب $f^*(x|\theta)$ 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 $H[f^*(x|\theta)]$ نزدیک باشد. بنابراین می‌توان شاخص ID را برای انجام مقایسه بین توزیعها براساس تابع کالبک-لیبلر به صورت

$$ID(f : f^*|\theta) = 1 - \exp[-K(f : f^*|\theta)] \\ = 1 - \exp\{H[f(x|\theta)] - H[f^*(x|\theta)]\}$$

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

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

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

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

و یک مقدار $ID(f : f^*|\theta)$ 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 $T_1(X) = X^2$ و $\theta_1 = \sigma^2$ باشد آنگاه توزیع ME در خانواده

$$\Omega_\sigma = \{f(x | \sigma^2) : E(X^2) = \sigma^2\}$$

با توجه به (۲) توزیع نرمال $N(0, \sigma^2)$ $f^*(x | \sigma^2) = N(0, \sigma^2)$ می باشد. خانواده های مقیاس-مکان زیر را در نظر بگیرید:
الف) لجستیک با چگالی

$$f_1(x | \theta_1, \lambda_1) = \lambda_1 \exp\{-\lambda_1(x - \theta_1)\} \times [1 + \exp\{-\lambda_1(x - \theta_1)\}]^2, \\ -\infty < \theta_1 < \infty, \lambda_1 > 0$$

ب) لاپلاس با چگالی

$$f_2(x | \theta_2, \lambda_2) = \frac{1}{2} \lambda_2 \exp\{-\lambda_2 |x - \theta_2|\}, \\ -\infty < \theta_2 < \infty, \lambda_2 > 0$$

ج) گمبل با چگالی

$$f_3(x | \theta_3, \lambda_3) = \lambda_3 \exp\{-\lambda_3(x - \theta_3)\} \times \exp[-\exp\{-\lambda_3(x - \theta_3)\}], \\ -\infty < \theta_3 < \infty, \lambda_3 > 0$$

θ_i و $\lambda_i, i = 1, 2, 3$ پارامترهای مکان و مقیاس توزیعها می باشند. برای

$$\pi^2/(3\lambda_1^2) = 2/\lambda_1^2 = \pi^2/(2\lambda_2^2) = \sigma^2$$

f_1 و f_2 اعضای Ω_σ می باشند. آنتروپی مقیاس مکان بالا با واریانس متناهی σ^2 به صورت

$$H[f_1(x | \sigma^2)] = \frac{1}{2} \ln(\sigma^2) + \frac{1}{2} \ln(3e^4/\pi^2) \\ H[f_2(x | \sigma^2)] = \frac{1}{2} \ln(\sigma^2) + \frac{1}{2} \ln(2e^2) \\ H[f_3(x | \sigma^2)] = \frac{1}{2} \ln(\sigma^2) + \frac{1}{2} \ln(7e^2/\pi^2) + \gamma$$

می باشد که $\gamma = 0.5771$ ثابت اولر (Euler Constant) می باشد.

با به کارگیری ماکزیمم آنتروپی یعنی

$$H[f^*(x | \sigma^2)] = \frac{1}{2} \ln(\sigma^2) + \frac{1}{2} \ln(2\pi e)$$

در عبارت

$$ID(f : f^* | \theta) = 1 - \exp\{H[f(x | \theta)] - H[f^*(x | \theta)]\}$$

شاخصهای ID خانواده های مقیاس-مکان بالا نسبت به نرمال بودن به صورت زیر بدست می آیند:

$$ID(f_1 : f^* | \theta) \approx 0.014$$

$$ID(f_2 : f^* | \theta) \approx 0.070$$

$$ID(f_3 : f^* | \theta) \approx 0.087$$

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

۴ آماره های ID

در بخش قبل شاخص ID را براساس تابع کالبد-لیبلر به صورت

$$ID(f : f^* | \theta) = 1 - \exp\{H[f(x | \theta)] - H[f^*(x | \theta)]\} \quad (5)$$

معرفی نمودیم. اکنون سعی می کنیم برآوردی از شاخص ID را ارائه دهیم. با توجه به عبارت (5) برآورد $ID(f : f^* | \theta)$ موکول به برآورد $H[f(x | \theta)]$ و $H[f^*(x | \theta)]$ می باشد. ماکزیمم آنتروپی به صورت:

$$H[f^*(x | \theta)] = -\ln M(\theta) - \lambda_1(\theta)\theta_1 - \dots - \lambda_m(\theta)\theta_m$$

تابعی از θ است. وقتی که پارامتر θ مقدار مشخص θ_0 باشد، $H[f^*(x | \theta)]$ مقدار معلومی خواهد بود. ولی در بیشتر مسائل، θ براساس داده ها برآورد شده و بنابراین برآورد $H[f^*(x | \theta)]$ به یک مسئله برآورد پارامتر تبدیل می شود. در صورت داشتن

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

$$ID_n(f : f^* | \theta_n) = 1 - \exp\{H_n[f] - H[f^*(x|\theta)_n]\}$$

که $0 \leq ID_n(f : f^* | \theta_n) \leq 1$ بوده و در صورتی که مقدار $ID_n(f : f^* | \theta_n)$ نزدیک صفر نباشد، باید قیود اضافی یا انواع دیگری از قیود را جستجو نمود [۵].

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

مراجع

- [1] Cover, T.M. and J.A. Thomas (1991) Elements of Information Theory. 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 Continuous Probability Distributions. IEEE. trans on Info. Theory IT- 24, 120-122.