دانشگاه فردوسی مشهد
دانشکده علوم ریاضی
گروه آمار

عنوان :

ارتباط بین برآوردگر آنتروپی ماکسیمم و برآوردگر درستنمایی ماکسیمم

مولف :

فاطمه عباس پور

ارائه شده جهت اخذ درجه کارشناسی ارشد در رشته آمار

استاد راهنما:

دکتر غلامرضا محتشمی برزادران

استاد مشاور:

دکتر عبدالحمید رضایی رکن آبادی

اسفند ۱۳۸۹

تقدیم به

پروردگار مهربانم که همواره ناظر بر اعمالم است

ساحت مقدس آقا علی بن موسی الرضا

پدر و مادر مهربان و بزرگوارم

و استاد گرانقدرم

قدردانی

حمد و سپاس پروردگاری را سزاست که به من توفیق تحصیل عنایت فرمود که اگر پرتو فضل او نبود هرگز موفق به انجام این مهم نمی‌شدم.
سپاس بی‌پایان خود را تقدیم استاد راهنمای بزرگوارم جناب آقای دکتر غلامرضا محتشمی برزادران می‌نمایم که بدون نظارت صبورانه ایشان در مراحل مختلف تحقیق و تدوین پایان نامه، به انجام رساندن این مجموعه غیرممکن بود. همچنین از استاد مشاور محترم آقای دکتر رضایی رکن آبادی کمال تشکر را دارم. از پدر عزیزم که‌همواره بهترین پشتیبان و راهنمایم بوده و مادر مهربانم که هیچگاه محبتهایش را از من دریغ نکرده و آنچه در توان داشت برای کسب تحصیل من انجام داده و همچنین از خواهران بسیار خوب و دلسوزم که همیشه مشوق من دراین راه بوده‌اند، بسیار متشکرم و امیدوارم بتوانم زحماتشان را جبران نمایم.
در نهایت از آقای یحیی محتشمی که زحمت بازنویسی برنامه ی نرم افزار $Matlab$ را به عهده داشتند، نیز سپاسگزاری می کنم و برایشان آرزوی موفقیت دارم.

فاطمه عباس پور— اسفند ۱۳۸۹

پیشگفتار

اندازه ی آنتروپی شانون به عنوان زیر بنای نظریه ی اطلاع نقش عمده ای در استنباط آماری دارد. این اندازه برای اولین بار توسط شانون در سال ۱۹۴۸ مطرح شده است. همچنین جینز در سال ۱۹۵۷ اصل آنتروپی ماکسیمم را ارائه نموده است. جعفری (۱۹۹۱)، گلان (۱۹۹۸) و گرندار(۲۰۰۰)، ارتباط بین برآوردگر درستنمایی ماکسیمم و برآوردگر آنتروپی ماکسیمم را مورد بررسی قرار داده اند. اندازه ی اطلاع کلبک و برآوردگر اطلاع کلبک مینیمم نیز در این پایان نامه مورد بررسی قرار گرفته اند. شور(۱۹۸۴)، اگوچی و کوپاس (۲۰۰۶) و شلنز(۲۰۰۷) نیز ارتباط بین برآوردگر درستنمایی ماکسیمم و برآوردگر اطلاع کلبک مینیمم را مطرح نموده اند.

در واقع هدف این پایان‌نامه، بررسی ارتباط بین برآوردگرهای ماکسیمم درستنمایی و ماکسیمم آنتروپی است و نیز یافتن ارتباط بین برآوردگرهای مینیمم اطلاع کلبک با برآوردگر ماکسیمم درستنمایی و همچنین یافتن برآورد ماکسیمم آنتروپی پارامترهای توزیع های مختلف به کمک برنامه ی نرم افزار ارائه شده است. این مجموعه شامل ۵ فصل می‌باشد که در زیر خلاصه‌ای از مطالب هرفصل آمده است.

- در فصل اول، مفاهیم و تعاریف مورد نیاز دراین پایان نامه، از جمله آنتروپی، اطلاع کلبک، تابع امتیاز و اطلاع فیشرشرح داده شده است.
- فصل دوم، شامل معرفی برآوردگر درستنمایی ماکسیمم و بیان برخی از ویژگی های آن و بیان ارتباط این برآوردگر با تابع امتیاز است.
- در فصل سوم، ابتدا آنتروپی و برخی از ویژگی های آن را معرفی می کنیم، اصل ماکسیمم آنتروپی رادر حالت پیوسته و گسسته مطرح می کنیم و به توصیف برخی از ویژگی های آنتروپی ماکسیمم می پردازیم.
- در فصل چهارم به بررسی ارتباط بین برآوردگر درستنمایی ماکسیمم و برآوردگر آنتروپی

ماکسیمم می پردازیم و به کمک برنامه ی نرم افزاری که ارائه شده برآورد آنتروپی ماکسیمم پارامترهای مدل ها و توزیع های مختلف را می یابیم.

- فصل پنجم نیز شامل معرفی برآوردگراطلاع کلبک مینیمم و بیان ارتباط آن با برآوردگر درستنمایی ماکسیمم و نیز معرفی برخی از اندازه های اطلاع دیگر و ارتباط برخی از آن ها با یکدیگر است.

علائم

$H(X)$: آنتروپی متغیر تصادفی گسسته ی X.

$h(X)$: آنتروپی متغیر تصادفی پیوسته ی X.

$H(X,Y)$: آنتروپی توأم متغیرهای تصادفی گسسته ی X و Y.

$H(X|Y)$: آنتروپی شرطی متغیر تصادفی گسسته ی X به شرط معلوم بودن Y.

H_{max}: ماکسیمم مقدار تابع آنتروپی در حالت گسسته وپیوسته .

ΔH: تفاضل ماکسیمم تابع آنتروپی از آنتروپی .

$D(p||q)$: اندازه ی اطلاع کلبک بین بردارهای احتمال p و q.

$I(X;Y)$: اطلاع متقابل یا اطلاع دوطرفه بین X و Y.

(X) : تابع امتیاز متغیر تصادفی X .

$I_X(\theta)$: میزان اطلاع فیشر درباره ی پارامتر θ در متغیر تصادفی X.

$I_{st}(X)$: اطلاع فیشر استاندارد شده ی X.

$D(X)$: اطلاع کلبک بین متغیر تصادفی X و متغیر تصادفی با توزیع نرمال دارای میانگین صفر و واریانس $\sigma^۲$.

$s(p)$: تابع تعجب برای تابع احتمال P.

$r(t)$: تابع نرخ خطر برای متغیر تصادفی t.

$L(\theta;x_۱,x_۲,...,x_n)$: تابع درستنمایی بردار $(x_۱,x_۲,...,x_n)$ بر حسب پارامتر θ.

$\ell(\theta)$: لگاریتم تابع درستنمایی .

D_ν: فاصله ی تفاضلی بین دو تابع احتمال دلخواه.

D_H: فاصله ی هلینگر بین دو تابع احتمال دلخواه.

$D_{\chi^۲}$: فاصله ی تفاضلی بین دو تابع احتمال دلخواه.

D_α: α فاصله ی بین دو تابع احتمال دلخوا.

D_B: فاصله ی باتاچاریا بین دو تابع احتمال دلخواه.

D_{Ha}: فاصله ی همساز بین دو تابع احتمال دلخواه.

D_J: فاصله ی جفریز بین دو تابع احتمال دلخواه.

$D_{\triangle}$: فاصله ی مثلثی بین دو تابع احتمال دلخواه.

D_f: فاصله ی سزار بین دو تابع احتمال دلخواه.

D_{LW}: فاصله ی لین و ونگ بین دو تابع احتمال دلخواه.

$L(p,q)$: امید ریاضی لگاریتم تابع احتمال q نسبت به تابع p.

Δ_{power}: فقدان توان روی تمام مقادیر ممکن .

$U(x)$:عامل کل .

$\tilde{U(x)}$:عامل نرمال شده .

$E^s(x)$: شدت اصلی متغیر X با تابع چگالی $f(x)$ روی تکیه گاه s.

$E^c(x)$: شدت سببی نسبت به عامل کلی $U(x)$.

H_{ME}: ماتریس هسین مشتقات دوم تابع لاگرانژ برای یافتن تابع با ماکسیمم مقدار آنتروپی تحت گشتاورهای داده شده.

H_{ML}: ماتریس هسین مشتقات دوم لگاریتم تابع درستنمایی نسبت به پارامتر.

$m(V)$: میانگین نمونه برای تابع $V(x)$.

$\mu(V)$: امیدریاضی تابع $V(x)$.

FOC: برقرار بودن شرایط سازگاری، یعنی شرایطی که تحت آن گشتاور نمونه با گشتاور جامعه برابر است .

فهرست جداول

فهرست مندرجات

فصل ۱

مفاهیم و مقدمات

۱.۱ مقدمه

در این فصل، مفاهیم و نمادهایی که در این پایان نامه مورد نیاز است را معرفی می کنیم. ابتدا مفاهیم نظریه ی اطلاع، آنتروپی، اطلاع کلبک، اطلاع متقابل، اطلاع فیشر و تابع امتیاز را مطرح نموده و سپس در قالب قضایا و مثال هایی ارتباط این مفاهیم را بیان می کنیم.

۲.۱ نظریه ی اطلاع

ارتباط به معنای ارسال اطلاعات از یک منبع از طریق یک کانال به یک مقصد است. حال این سوال مطرح می شود که اطلاع چیست ؟ اندازه گیری بعضی چیزها ساده است. حداقل در اصل، اندازه گیری بعضی چیزها قابل تصور و ادراکی است مانند وزن، فاصله یا ارتفاع.
اما اطلاع را چگونه می توان اندازه گیری کرد؟ نظریه ی اطلاع می تواند یک پاسخ ساده برای این سوال فراهم کند.
پیشرفت این رشته، مدیون شانون [۱] است. او در سال ۱۹۴۸ یک چارچوب برای مسائل مورد مطالعه ی انتقال اطلاعات چاپ کرد.شانون معتقد بود که اطلاع عبارت است از اینکه به یک احتمال اجازه ی متمایز بودن در میان دیگر احتمالات رابدهیم. این تعریف را می توانیم در نظریه ی مخابرات نیز به کار ببریم: یک پیام از یک منبع اولیه به یک مقصد فرستاده می شود، این پیام به دریافت کننده اجازه می دهد تا میان دیگران متمایز باشد و به این ترتیب اطلاع انتقال می یابد. حال می خواهیم بدانیم که اطلاع چگونه اندازه گیری می شود. نظریه ی اطلاع به دو سوال اساسی در نظریه ی مخابرات پاسخ می دهد.
۱– فشرده سازی نهایی داده ها چیست؟
که پاسخ آن آنتروپی است.
۲– نرخ انتقال ارتباط نهایی چیست؟

---
[۱] Shannon

و پاسخ آن ظرفیت کانال است.

به همین دلیل است که عده ای نظریه ی اطلاع را زیر مجموعه ی نظریه ی مخابرات می دانند.

امروزه، نظریه ی اطلاع را نقطه ی فرین نظریه ی مخابرات می دانند .

نظریه ی اطلاع را معمولاً به عنوان محصول زمان جنگ جهانی دوم و بیشتر آن تلاش و فعالیت های شانون می دانند. شانون در آوریل ۱۹۱٦ در میشیگان متولد شد وی در سال ۱۹۴۰ درجه ی استادی خود را در رشته های مهندسی الکترونیک و ریاضیات دریافت کرد. در سال ۱۹۴۱ به عنوان یک مهندس مخابرات برای شرکت سیستم بل شروع به فعالیت کرد. وی در طول جنگ جهانی دوم در گروهی در شرکت بل به فرآیند کدسازی پیام های میان رزولت و چرچیل به منظور حفظ امنیت اطلاعات مبادله شده می پرداخت.

وی دیدگاهی در مورد حد بالایی کارایی کدسازی که ما را به نظریه ی ریاضی مخابرات رهنمون می کند، به دست آورد. در واقع او شاخه ی جدیدی از علم را به وجود آورد. شانون اندازه ی آنتروپی متغیر تصادفی X با بردار احتمال $P = (p_۱, p_۲, ..., p_n)$ را به صورت $H(X) = -\sum_{i=۱}^{n} p_i \log p_i$ تعریف می کند که در آن $\sum_{i=۱}^{n} p_i = ۱$ است. و پس از او به پاس تلاش های وی نام این اندازه را آنتروپی شانون نامیدند.

متن اصلی مقاله ی شانون در سال ۱۹۴۸ در مجله ی تکنیکی سیستم بل منتشر شد وبعد از آن در یک کتاب با عنوان نظریه ی ریاضی ارتباطات به چاپ رسید. شانون در ۲۴ فوریه ی ۲۰۰۱ در سن ۸۴ سالگی فوت کرد.

سیستم های ارتباطی موجود تا قبل از سال ۱۹۴۸ را می توان به صورت زیر نام برد:

- تلگراف (۱۸۳۰)، تلفن (۱۸۷٦).

- تلگراف بی سیم (۱۸۸۷).
- رادیوی AM (۱۹۰۰).
- تلویزیون (۱۹۲۷—۱۹۲۵).
- پخش کننده ی بینایی (۱۹۴۰).

در سال ۱۹۲۸، نیکویست [۲] مطلبی را منتشر کرد که در آن به این موضوع پرداخته بود که چگونه‌می تواند یک پیام در یک کانال تلگراف با حداکثر سرعت ممکن و بدون تحریف فرستاده شود.
هارتلی [۳](۱۹۲۸)، تلاش کرد تا یک اندازه از اطلاع را به عنوان لگاریتم تعداد پیام های قابل تمایز معرفی کند. شانون نیز اندازه ی اطلاعی را معرفی کرد که بر اساس پذیره ی احتمالی است که شامل اندازه ی هارتلی به عنوان یک نمونه ی خاص است.
در ابتدای دهه ی ۱۹۴۰، فرستادن اطلاعات به عنوان یک نرخ مثبت با احتمال خطای جزئی غیرممکن به نظر می رسید. در سال ۱۹۴۸ در نظریه ی مخابرات احساس شد که منبع اطلاع باید به عنوان یک فرآیند تصادفی در نظر گرفته شود. اما شانون توانست احتمال خطای موجود در ظرفیت کانال را به نزدیک صفر برساند. شانون معتقد بود که اگر آنتروپی منبع کمتر از ظرفیت کانال باشد، خطای ارتباط به طور مجانبی می تواند به حداقل مقدار خود برسد. در طول ۶۰ سال اخیر، افرادی مانندکلبک [۴] (۱۹۵۹) و کاور و توماس [۵](۲۰۰۶)، رنی [۶](۱۹۶۱) و جینز[۷](۱۹۵۷) به تکمیل و بررسی مقاله ها و کارهای شانون پرداخته اند.
ماکسیمم سازی آنتروپی در کلاسی از توزیع ها، تحت قیدهایی معین، نقشی مهم در نظریه های آماری دارد.

---

[۲] Nyquist
[۳] Hartley
[۴] Kullback
[۵] Cover and Thomas
[۶] Renyi
[۷] Jaynes

در بیشتر کتاب های نظریه ی اطلاع، می توانیم مراجع زیادی را بیابیم که منجر به تحول در نظریه ی اطلاع شده اند. در ادامه به کمک دو مثال با آنتروپی بیشتر آشنا می شویم.

مثال ۱.۱ متغیرهای تصادفی X و Y را با احتمالات $P = (۰/۸, ۰/۲)$ و $Q = (۰/۲, ۰/۸)$ در نظر بگیرید. در این صورت $H(X) = H(Y) = ۰/۷۲$ می باشد. بنابراین می توان چنین نتیجه گرفت که مقدار احتمالات در مقدار آنتروپی مهم است نه ترتیب قرار گرفتن آن ها.

مثال ۲.۱ گاهی اوقات نیز ممکن است احتمالات متفاوت، به مقدار آنتروپی یکسانی منجر شود. برای مثال فرض کنید که‌متغیرهای X و Y با احتمالات $P = (۰/۵, ۰/۲۵, ۰/۲۵)$ و $Q = (۰/۴۸, ۰/۳۲, ۰/۲)$ داده شده باشند، در این صورت $H(X) = H(Y) = ۱۵$ است.

۳.۱ چرا آنتروپی؟

اگر X یک متغیر تصادفی با تابع احتمال p باشد، می توانیم تابع آنتروپی را به صورت امید ریاضی تابع تعجب [۸] تعریف کنیم . اما تابع تعجب را به صورت $S(p) = -\log_a p$ تعریف می کنیم که p در آن احتمال پیشامد مورد نظر است. حال سوالی که مطرح می شود این است که چرا چنین تابع تعجبی را برای تعریف آنتروپی انتخاب کرده ایم؟ در پاسخ به این سوال باید گفت که این تابع باید در شروط زیر صدق کند.

شرط ۱) $s(۱) = ۰$ (میزان تعجب پیشامد حتمی صفر است).

شرط ۲) اگر $p < q$ باشد آنگاه $s(p) > s(q)$ باشد. یعنی s تابعی نزولی اکید از p باشد(میزان تعجب پیشامدی که شانس کمتری دارد، بیشتر است).

شرط ۳) $s(p)$ تابعی پیوسته از p باشد.

---

[۸] Surprise function

شرط ۴) $s(pq) = s(p) + s(q)$ باشد، به ازای هر p و q متعلق به فضای $[0, 1)$(در حقیقت میزان تعجب دو پیشامد مستقل برابر مجموع میزان تعجب آن دو پیشامد است).
به کمک این شرایط فرم تابعی $s(p)$ به صورت زیر است:
از شرط ۴ داریم:

$$s(p^2) = 2s(p),$$

و به کمک استقرا داریم:

$$s(p^m) = ms(p), \quad m \in Z,$$

و برای مقدار صحیح n نتیجه می شود:

$$s(p) = s(p^{\frac{1}{n}}.....p^{\frac{1}{n}}) = ns(p^{\frac{1}{n}}),$$

$$s(p^{\frac{1}{n}}) = \frac{1}{n}s(p),$$

$$s(p^{\frac{m}{n}}) = \frac{m}{n}s(p),$$

که معادل است با $s(p^x) = xs(p)$ که در آن x مقداری گویا و مثبت است.
اما شرط سوم، پیوستگی تابع s بود. پیوستگی تعاریف متعددی دارد. یکی از تعاریف پیوستگی این است که تابع s پیوسته است اگر دنباله ی q_n به q همگرا باشد، تابع $s(q_n)$ هم به $s(q)$ همگرا باشد. حال فرض می کنیم که $q = p^x$ باشد که در آن x مقداری دلخواه و مثبت است و نیز $q_n = p^{\frac{[nx]}{n}}$ که منظور از $[nx]$ ، جزٔ صحیح nx است. بنابراین $\lim_{n\to\infty} q_n = q = p^x$ است و چون

$$s(q_n) = s(p^{\frac{[nx]}{n}}) = \frac{[nx]}{n}s(p),$$