بسم الله الرحمن الرحیم

دانشگاه یزد

دانشکده ریاضی

گروه آمار

پایان‌نامه

برای دریافت درجه کارشناسی ارشد
آمار

مشخص‌سازی توزیع براساس آماره‌های ترتیبی ورکوردها

استاد راهنما: دکتر حجت اله ذاکرزاده

استاد مشاور: دکتر عیسی محمودی

پژوهش و نگارش: سارا جوادی

مهر ماه ۱۳۹۱

تقدیم به مهربان فرشتگانی که لحظات ناب باور بودن، لذت و غرور دانستن، جسارت خواستن، عظمت رسیدن و تمام تجربه‌های یکتا و زیبای زندگیم، مدیون حضور سبز آنهاست، تقدیم به خانواده عزیزم.

شکر و سپاس خدا را که بزرگترین امید و یاور در لحظه لحظه زندگیست. بر خود واجب می‌دانم از زحمات استاد راهنمای بزرگوارم، جناب آقای دکتر ذاکرزاده که با راهنمایی‌های فراوانشان، مرا یاری نمودند، تشکر نمایم. همچنین از خانواده‌ی عزیزم که پیوسته یاریگرم بودند و هر لحظه تلاشم با فداکاری آن‌ها میسر گشته، سپاس‌گزاری می‌نمایم.

چکیده

در این پایان‌نامه تعدادی از توزیع‌های پیوسته، براساس امید شرطی تابعی از مقادیر رکورد و خصوصیات امیدهای شرطی آماره‌های ترتیبی سانسور شده‌ی پیش‌رونده نوع دوم مشخص می‌شوند، توزیع‌های مختلف ممکن است آنتروپی رنی یکسانی داشته باشند، بنابراین یک توزیع به‌طور منحصربه‌فرد، به‌وسیله آنتروپی رنی مشخص نمی‌شود، در این متن خصوصیات آنتروپی رنی آماره‌های ترتیبی را توصیف نموده و تعدادی از مشخص‌سازی‌ها را بر اساس آنتروپی رنی آماره‌های ترتیبی و رکوردها مورد بررسی قرار گرفته و همچنین برخی مشخص‌سازی‌ها بر اساس آنتروپی شانون این آماره‌ها انجام شده است.

لیست جداول

پیشگفتار

در مباحث آماری توابع مختلفی جهت مشخص‌سازی توزیع یک متغیر تصادفی وجود دارد، از جمله توابع مهمی که می‌توان توسط آن‌ها توزیع را تشخیص داد، تابع میانگین از راست سانسور شده است، برخی از آماردانان نیز موضوع مشخص‌سازی F را بر اساس خصوصیات آماره‌های ترتیبی و مقادیر رکورد مورد مطالعه قرار دادند، مقالات هانگ[۱] [۲۹] در سال ۱۹۷۵، ناگاراجا[۲] [۳۳] در سال ۱۹۸۸، ناگاراجا و نوزروف[۳] [۳۳] در سال ۱۹۹۷، شامل مشخص‌سازی‌هایی براساس آماره‌های ترتیبی و مقادیر رکورد است. احسان‌اله[۴] [۸] در سال ۱۹۸۲ توزیع نمایی را براساس برخی خصوصیات مقادیر رکورد مشخص نمود. همچنین او در سال ۱۹۹۲ برخی خصوصیات مقادیر رکورد را از توزیع نمایی مورد بررسی قرار داد. همچنین احسن‌اله و بالاکریشنان[۵] [۱۵] مطالعاتی را در زمینه‌ی گشتاورهای مقادیر رکورد از توزیع پارتوی تعمیم یافته در سال ۱۹۹۴ انجام داده‌اند. سپس احسان اله [۸] نیز در سال ۱۹۹۵ روابطی را برای گشتاورهای مقادیر رکورد از توزیع نمایی ارائه نمود. در سال ۱۹۹۶ رویز و ناوارو [۶] [۳۶] نمایش جالبی از $F(x)$ را برپایه‌ی تابع میانگین ناقص دوتایی ارائه دادند. بالاکریشنان و همکارانش در همان سال ارتباط بین آماره‌های ترتیبی و تابع میانگین دوتایی ناقص را به‌صورت زیر نشان دادند،

$$m(x,y)=E(X|x\leq X\leq y)=\frac{1}{F(y)-F(x)}\int_{x}^{y} tdF(t),$$

به‌طوری‌که دامنه‌ی تعریف $m(x,y)$ به‌صورت $D=\{(x,y)\in R^{2}:F(x)<F(y)\}$ است. آماره‌های ترتیبی در دامنه‌ی وسیعی از مسائل که شامل مشخص‌سازی توزیع‌های احتمال، آزمون‌های نیکویی برازش، برآورد آنتروپی، تحلیل نمونه‌های سانسور شده، تحلیل‌های قابلیت اعتماد و کنترل کیفیت کالاها می‌شود، مورد استفاده قرار می‌گیرند. در نظریه‌ی قابلیت اعتماد، آماره‌های ترتیبی و مقادیر رکورد در مدل‌بندی‌های آماری به‌کاربرده می‌شوند. احمدی و بالاکریشنان [۴] در سال ۲۰۰۵، فواصل اطمینان را برای چارک‌ها

---

[۱] Hung

[۲] Nagaraja

[۳] Nevzorov

[۴] Ahsanullah

[۵] Balakrishnan

[۶] Ruiz and Navarro

براساس رکوردهای جاری و دامنه‌ی رکوردها به‌دست آوردند. برات‌پور و همکارانش [۲۰] در سال ۲۰۰۷ درباره‌ی خصوصیات اطلاع شانون از آنتروپی رکوردها فعالیت‌هایی انجام داده‌اند.

در این پژوهش نیز مشخص‌سازی توزیع بر اساس مقادیر رکورد و آماره‌های ترتیبی مورد مطالعه قرار گرفته است، فصول مختلف این پایان‌نامه به شرح زیر است:

در فصل اول مفاهیم و تعاریف اولیه‌ی مورد نیاز این پایان‌نامه آورده شده‌اند. در فصل دوم مشخص‌سازی بر پایه‌ی آماره‌های ترتیبی و تابع میانگین محدب شرطی طبق شرایط خاصی بررسی شده است، همچنین در این فصل به‌صورت خاص مشخص‌سازی بر اساس آماره‌های ترتیبی سانسور شده‌ی پیش‌رونده نوع دوم بررسی شده و در پایان فصل با توجه به مفهوم آنتروپی رنی و آماره ترتیبی، سیستم‌های سری، موازی و m از n مشخص‌سازی شده‌اند. در فصل سوم این پایان‌نامه مشخص‌سازی براساس رکوردها مورد مطالعه قرار گرفته است. در ابتدای این فصل، مسئله‌ی مشخص‌سازی متغیر تصادفی X را به کمک امیدهای شرطی تابعی از مقادیر رکورد بررسی نموده، در ادامه مشخص‌سازی بر اساس آنتروپی رنی مقادیر رکورد و آنتروپی شانون رکوردهای جاری مورد مطالعه قرار گرفته است. در پایان فصل سوم نیز چند خصوصیت آنتروپی شانون رکوردهای جاری آورده شده‌اند.

فصل ۱

مفاهیم و تعاریف اولیه

مقدمه

در این فصل به بیان مفاهیم اولیه مانند مشخص‌سازی، آنتروپی شانون، آنتروپی رنی، ویژگی‌های آنتروپی رنی، رکورد، رکورد جاری، انواع سانسور و مفاهیم اولیه‌ی آنالیز ریاضی و دیگر تعاریف استفاده شده در این پایان‌نامه پرداخته شده است. همچنین چند مفهوم از ترتیب‌های تصادفی در متغیرهای تصادفی مطرح شده است.

۱.۱ مفاهیم آماری

تعریف ۱.۱.۱ (مشخص‌سازی) منظور از مشخص‌سازی، تعیین شرایطی است که شامل خصوصیات خاصی از متغیر تصادفی X است به‌طوری که بتوان با کمک آن شرایط، تابع توزیع مربوطه را تعیین کرد. در واقع منظور ویژگی‌ای است که به‌صورت منحصر به‌فرد، تابع توزیع را بر اساس توابعی از متغیرهای تصادفی که توزیع آن‌ها به X مربوط است، مشخص می‌کند.

تعریف ۲.۱.۱ (آنتروپی رنی)

مفهوم آنتروپی اولین بار در علم فیزیک معرفی و برای کمی‌سازی مفاهیمی چون عدم حتمیت و بی‌نظمی به‌کار برده شد. بعدها از این مفهوم برای اندازه‌گیری اطلاعات حاصل از مشاهدات متغیرهای تصادفی استفاده شد. اگر X یک متغیر تصادفی با تابع چگالی f و تابع توزیع F باشد، آن‌گاه آنتروپی مرتبه α یا آنتروپی رنی یک توزیع، به‌صورت زیر تعریف می‌شود:

$$\begin{aligned} H_\alpha(X) &= \frac{۱}{۱-\alpha} \log \int_{-\infty}^{+\infty} f^\alpha(x)dx \\ &= \frac{۱}{۱-\alpha} \log \ E[f_X^{\alpha-۱}(x)], \end{aligned}$$

که در آن $\alpha > ۰$، $\alpha \neq ۱$.

نکته ۳.۱.۱ خصوصیات آنتروپی رنی درخانواده‌های مکان و مقیاس به‌صورت زیر می‌باشند.

آنتروپی رنی در حالت مکانی مستقل از θ است و برای حالتی از خانواده‌ی مقیاس تابعی از $\log\ \theta -$ است. برای مثال همان‌طور که در جدول دیده می‌شود، چون توزیع نمایی از خانواده چگالی مقیاس است، آنتروپی رنی آن تابعی از $\log\ \theta -$ شده است. در جدول ۱.۱ آنتروپی رنی چند توزیع مشهور ارائه شده است.

جدول ۱.۱: آنتروپی رنی توزیع‌های مشهور

توزیع	$f(x)$	$H_\alpha(X)$
نمایی	$\lambda e^{-\lambda x}$	$-\log\lambda - \frac{1}{1-\alpha}\log\alpha$
پارتو	$\lambda\beta^{\lambda}x^{-\lambda+1}$	$\log\beta + \frac{\alpha}{1-\alpha}\log\lambda - \frac{1}{1-\alpha}\big[(1+\lambda)\alpha - 1\big]$
نرمال	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\big[-\frac{1}{2\sigma^2}(x-\mu)^2\big]$	$\log\sigma + \frac{1}{2}\log 2\pi - \frac{1}{2(1-\alpha)}\log\alpha$
وایبل	$\beta\lambda^{\beta}x^{\beta-1}\exp\big[-(\lambda x)^{\beta}\big]$	$-\log\lambda + \frac{1}{1-\alpha}\log\left[\frac{\Gamma(\alpha-\frac{\alpha-1}{\beta})}{\alpha^{\alpha}}\right] - \log(\beta\alpha)^{\beta}$
بتا	$\frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)}$	$(1-\alpha)^{-1}\log\left[\frac{\beta\big(\alpha(a-1)+1,\alpha(b-1)+1\big)}{\beta^{\alpha}(a,b)}\right]$

(الف) خانواده‌ی چگالی مکان: μ حقیقی و $F_\mu(x) = F(x-\mu)$.

(ب) خانواده‌ی چگالی مقیاس: $\sigma > 0$ و $F_\sigma(x) = F(\frac{x}{\sigma})$.

تعریف ۴.۱.۱ (آنتروپی شانون)

برای تابع چگالی احتمال پیوسته $f(x)$ روی مجموعه S، شانون، آنتروپی متغیر تصادفی را به‌صورت زیر تعریف کرد:

$$H(X) = -\int_S f(x)\log\ f(x)dx, \tag{1.1}$$

و با استفاده از تبدیل انتگرال احتمال داریم:

$$H(X) = -\int_0^1 \log f(F^{-1}(x))dx,$$

برای متغیر تصادفی گسسته X روی مجموعه $S = \{x_1, x_2, ...\}$ با تابع جرم احتمال $\{p_1, ..., p_n\}$ که $p_i = p(x_i)$ است، آنتروپی شانون به‌صورت زیر تعریف می‌شود.

$$H(p) = -\sum_{i\geq 1} p_i log\ p_i.$$

در چگالی‌های پیوسته، نمادهای دیگری از جمله $H(X)$ و $H(F)$ نیز برای آنتروپی استفاده می‌شود. از نظر فیزیکی، سیستم‌ها باید با حالت‌هایی با بیشترین آنتروپی ظاهر شوند تا به تعادل برسند. در نظریه‌ی احتمال به $H(p)$، به‌عنوان یک اندازه اطلاعی که p دارد نگاه می‌کنند. بنابراین آنتروپی بیشتر به معنای بیشتر بودن کمبود اطلاع در مورد یک پدیده است.

مثال ۵.۱.۱ مجموعه متناهی $\{x_۱, x_۲, ..., x_n\}$ را در نظر بگیرید. اگر $p(x_۱) = ۱$ و برای $j > ۱$، $p(x_j) = ۰$ آن‌گاه

$$H(p) = -۱ \log \ ۱ = ۰.$$

تابع آماری به وسیله p تنها یک برآمد ممکن $x_۱$ را ارائه می‌دهد و آگاهی کامل از آنچه اتفاق خواهد افتاد داریم. از طرف دیگر اگر p تابع چگالی احتمال یکنواخت گسسته باشد که در آن برای همه jها $p(x_j) = \frac{۱}{n}$، آن‌گاه $H(p) = \log \ n$. خواهیم دید که هر تابع چگالی احتمال روی $\{x_۱, ..., x_n\}$ مقدار آنتروپی کوچکتر یا مساوی با $\log \ n$ دارد و مقدار آنتروپی $\log \ n$ فقط برای توزیع یکنواخت رخ می‌دهد. به بیان دیگر تابع چگالی احتمال روی $\{x_۱, ..., x_n\}$ با بیشینه‌ی آنتروپی، معادل با کمترین مقدار آگاهی از $\{x_۱, ...x_n\}$ است.

در مثال‌های زیر مقدار آنتروپی شانون برای توزیع نرمال و نمایی محاسبه شده است.

مثال ۶.۱.۱ آنتروپی چگالی نرمال روی $\mathbb{R}$ با میانگین μ و واریانس $\sigma^۲$ عبارت است از:

$$\begin{aligned} H(F) &= -\int_{\mathbb{R}} \frac{۱}{\sqrt{۲\pi\sigma^۲}} e^{-\frac{(x-\mu)^۲}{۲\sigma^۲}} (-\log \ (\sqrt{۲\pi\sigma^۲}) - \frac{۱}{۲}(\frac{x-\mu}{\sigma})^۲) dx \\ &= \frac{۱}{۲}(۱ + \log \ (۲\pi\sigma^۲)). \end{aligned} \tag{۲.۱}$$

دیده می‌شود که μ در میزان آنتروپی نقشی ندارد. پس تمام توزیع‌های نرمال با واریانس $\sigma^۲$ دارای یک آنتروپی مشابه هستند. برای $\sigma^۲$ های نزدیک به صفر ($\sigma^۲ \leqslant \frac{e^{-۱}}{۲\pi}$)، مقدار آنتروپی منفی است.

از نظر نموداری وقتی $\sigma^۲$ کوچک است، بخش قابل توجهی از تابع چگالی احتمال مقداری بیشتر، از یک را در بر می‌گیرد و در نتیجه $-f log \ f < ۰$. برای توزیع‌های گسسته آنتروپی همواره بزرگتر از صفر است زیرا مقادیر یک تابع چگالی احتمال گسسته هرگز از مقدار یک تجاوز نمی‌کند. آنتروپی منفی در موارد پیوسته بازتاب این امر است که توزیع احتمال پیوسته می‌تواند متمرکزتر از یک توزیع یکنواخت روی $[۰, ۱]$ باشد.

مثال ۷.۱.۱ آنتروپی شانون متغیر تصادفی از توزیع نمایی با میانگین $\frac{۱}{\lambda}$ عبارت است از

$$H(F) = -\int_۰^\infty \frac{۱}{\lambda} e^{-\frac{x}{\lambda}} (-\log \ \lambda - \frac{x}{\lambda}) dx = ۱ + \log \ \lambda,$$

که مانند مثال ۶.۱.۱ در اینجا نیز آنتروپی برای λهای کوچک ($\lambda < e^{-1}$) می‌تواند منفی شود.

نکته ۸.۱.۱ (ویژگی‌های آنتروپی شانون)

آنتروپی شانون در چهار شرط زیر که به‌عنوان ویژگی‌های آن مطرح می‌شود صدق می‌کند.

۱) تابع $H(p)$ برحسب p پیوسته است.

۲) تابع $H(p)$ تعویض‌پذیر است، یعنی ترتیب احتمال‌های $p_1, \dots, p_n$ بر مقدار $H(p)$ تاثیری ندارد.

۳) تابع $H(p)$ جمع‌پذیر است، به‌عنوان مثال فرض کنید $H(X)$ مقدار آنتروپی مربوط به پرتاب یک تاس و $H(Y)$ مقدار آنتروپی مربوط به پرتاب تاس دیگر باشد، توجه کنید در این حالت $H(X) = H(Y)$ در حالی که $H(X,Y)$ مربوط به پرتاب دو تاس همزمان است، آن‌گاه در این مثال نتیجه می‌شود که $H(X,Y) = H(X) + H(Y)$.

۴) اگر تمام احتمال‌ها مساوی باشند $H(p)$ بیشینه می‌شود. این مطلب با حالتی متناظر است که در آن عدم قطعیت بیشینه وجود دارد.

نکته ۹.۱.۱ برای متغیر تصادفی توام (X,Y) با فضای مقادیر (M,N) و تابع احتمال توأم $p(x,y)$، میزان آنتروپی توام به‌صورت

$$H(X,Y) = -\sum_{x\in M}\sum_{y\in N} p(x,y)\log\ p(x,y),$$

تعریف می‌شود که آن را می‌توان به‌صورت زیر بیان کرد:

$$H(X,Y) = -E(\log\ p(X,Y)).$$

تعریف ۱۰.۱.۱ اگر (X,Y) دارای تابع چگالی احتمال توأم $p(x,y)$ باشد، آن‌گاه اندازه آنتروپی شرطی $H(Y|X)$ به‌صورت زیر تعریف می‌شود:

$$\begin{aligned} H(Y|X) &= \sum_{x\in M} p(x) H(Y|X=x) \\ &= -\sum_{x\in M} p(x) E_{Y|X}(\log\ p(Y|X=x)) \\ &= -\sum_{x\in M} p(x) \sum_{y\in N} p(y|x)\log\ p(y|x) \end{aligned}$$

$$= -\sum_{x\in M}\sum_{y\in N} p(x,y)\log\ p(y|x)$$
$$= -E_{X,Y}[\log\ p(Y|X)].$$

قضیه ۱۱.۱.۱ اگر X و Y مستقل باشند، داریم:

$$H(X,Y) = H(X) + H(Y) \tag{۳.۱}$$

و اگر X و Y وابسته باشند،

$$H(X,Y) = H(X) + H(Y|X) \tag{۴.۱}$$

اثبات: برای اثبات ۳.۱ با توجه به تعریف تابع H داریم:

$$H(X,Y) = -\sum_{x\in M}\sum_{y\in N} p(x,y)\log\ p(x,y)$$
$$= -\sum_{x\in M}\sum_{y\in N} p(x)p(y)(\log\ p(x) + \log\ p(y))$$
$$= -\sum_{x\in M} p(x)\log\ p(x)\sum_{y\in M} p(y) - \sum_{y\in N} p(y)\log\ p(y)\sum_{x\in M} p(x)$$

$$= H(X) + H(Y). \tag{۵.۱}$$

برابری آخر با توجه به رابطه زیر به‌دست می‌آید.

$$\sum_{y\in N} p(y) = \sum_{x\in M} p(x) = ۱.$$

در صورتی که X و Y وابسته باشند،

$$H(X,Y) = -\sum_{x\in M}\sum_{y\in N} p(x,y)\log\ p(x,y)$$
$$= -\sum_{x\in M}\sum_{y\in N} p(x,y)(\log\ p(x)p(y|x))$$
$$= -\sum_{x\in M}\sum_{y\in N} p(x,y)\log\ p(x) - \sum_{x\in M}\sum_{y\in N} p(x,y)\log\ p(y|x).$$

$$= H(X) + H(Y|X). \tag{۶.۱}$$