

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی شاهرود

دانشکده: ریاضی

گروه: آمار

مباحثی در نظریه اطلاع بر اساس داده های رکوردی

دانشجو:

حمید حسین خانی

اساتید راهنما:

دکتر محمد آرشی دکتر محسن مددی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ماه ۱۳۹۱

تعهد نامه

اینجانب حمید حسین خانی دانشجوی دوره کارشناسی ارشد رشته آمار گرایش ریاضی دانشکده ریاضی دانشگاه صنعتی شاهرود نویسنده پایان نامه مباحثی در نظریه اطلاع بر اساس داده های رکوردی تحت راهنمایی دکتر محمد آرشی و دکتر محسن مددی متعهد می شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج مه جود د، پایان نامه بدون ذکر منبع مجاز نمی باشد.

تقدیم به

آستان حقیقت و آمان که خود، مین حقیقت اند.

خدای رابی شاکرم که از روی کرم خانواده ای فداکار نصیم ساخت تا در سایه ی درخت پر بار

و جودشان بیایم و از ریشه آنها شاخ و برگ گیرم و از سایه و جودشان در راه کسب علم و دانش

تلاش نمایم. خانواده ای که بودنشان تاج افتخاری است بر سرم و نشان دلیلی است بر بودنم.

تقدیر و شکر

بر حسب وظیفه از اساتید راهنمای ارجمندم آقایان دکتر محسن مددی و دکتر محمد آرشی که زحمات فراوانی جهت هدایت، راهنمایی و همیاری بنده کشیده‌اند، تشکر می‌نمایم، بی‌شک این تحقیق توفیق نمی‌یافت مگر با راهنمایی‌های مشفقانه و بی‌دریغ ایشان. از خانم مریم نقوی که نظرات ارزنده‌ای در هر چه بهتر شدن این تحقیق داشتند، خاضعانه سپاسگزارم.

همچنین از دوست عزیزم آقای مهدی صالح نیا دانشجوی مقطع کارشناسی ارشد آمار دانشگاه شهید باهنر کرمان، بمنظور مشاوره‌ها و ارائه‌ی رهنمودهای مفید و سازنده کمال قدردانی را می‌نمایم.

در پایان از پدر و مادر عزیزم که با تحمل دشواری‌ها، سبب شدند تا آموختن در من زنده بماند صمیمانه سپاسگزارم.

چکیده پایان نامه

اخیراً چند محقق آنتروپی شانون (اطلاع شانون) را در داده‌های رکورد بررسی کرده‌اند، مانند برات پور و زاهدی (۲۰۰۷)، این آنتروپی را می‌توان با موفقیت برای کمی‌سازی میزان اطلاع (در رابطه با خانواده توزیع‌ها) استفاده کرد. مددی و تاتا در سال ۲۰۱۱ در یک دنباله از متغیرهای تصادفی پیوسته (iid)، روابطی برای اطلاع شانون ماکزیمم رکورد و مقادیر و زمان رخداد رکوردها در مدل کلاسیک (یک نمونه تصادفی در اندازه n) و طرح نمونه‌گیری معکوس به دست آورده‌اند، که می‌توان این اطلاع را برای توزیع‌های مشهور آماری محاسبه نمود. در این رساله به بررسی اطلاع شانون مقادیر رکوردهای بالا و پایین و زمان رخداد رکوردها برای توزیع مشهور t -استودنت می‌پردازیم. این رساله مشتمل بر ۳ فصل می‌باشد. در فصل ۱، تاریخچه رکوردها، آنتروپی و برخی از تعاریف و کاربردهای مقادیر رکورد مطرح شده‌اند. در فصل ۲، به ارایه نتایج اطلاع شانون مقادیر رکورد بالا و پایین و ارتباط زمان رخداد رکوردها در مدل کلاسیک و طرح نمونه‌گیری معکوس می‌پردازیم. در فصل ۳، به محاسبه اطلاع شانون مقادیر و زمان رخداد رکوردها در توزیع t -استودنت می‌پردازیم. در پیوست برنامه‌ها و دستورهای نوشته شده در نرم افزار *MAPLE* ارایه شده‌اند. در این پایان‌نامه، قضیه‌ها و نتایجی که صورت آن‌ها موجود و برهان توسط نویسنده پایان‌نامه می‌باشد، با علامت (*) و مواردی که، هم قضیه و هم برهان جدید می‌باشند، با علامت (**) مشخص شده‌اند.

حسین خانی

بهمن ماه ۱۳۹۱

نمادها

SI : اطلاع شانون

ISP : طرح نمونه‌گیری معکوس

$H_{RT}^L(n), H_{RT}^U(n)$: اطلاع شانون مقادیر و زمان رکوردهای بالایی و پایینی در مدل کلاسیک

$H_R^L(n), H_R^U(n)$: اطلاع شانون مقادیر رکوردهای بالایی و پایینی در مدل کلاسیک

$H_M^L(n), H_M^U(n)$: اطلاع شانون آخرین رکوردهای بالایی و پایینی در مدل کلاسیک

$h_{RT}^L(m), h_{RT}^U(m)$: اطلاع شانون مقادیر و زمان رکوردهای بالایی و پایینی در طرح نمونه‌گیری

معکوس

$h_R^L(m), h_R^U(m)$: اطلاع شانون مقادیر رکوردهای بالایی و پایینی در طرح نمونه‌گیری معکوس

$h_M^L(m), h_M^U(m)$: اطلاع شانون آخرین رکوردهای بالایی و پایینی در طرح نمونه‌گیری معکوس

γ : ثابت اویلر

$\psi(\cdot)$: تابع دی‌گاما

$\zeta(\cdot)$: تابع زتا

$t_{(r)}$: توزیع t - استودنت با درجه آزادی r

فهرست مندرجات

فصل ۱ مفاهیم مقدماتی..... ۱

۱-۱- مقدمه ۲

۱-۱-۱- مدل کلاسیک ۹

۱-۱-۲- طرح نمونه گیری معکوس 10

۱-۲- آنتروپی 11

۱-۳- معایب و مزایای استفاده از اطلاع شانون به جای عدم قطعیت 18

فصل ۲ اطلاع در داده های رکوردی ۱۹

۲-۱- اطلاع شانون در مقادیر رکورد ۲۰

۲-۱-۱- اطلاع در یک نمونه محدود از مقادیر رکوردها ۲۰

۲-۱-۲- اطلاع در مقادیر رکورد از یک طرح نمونه گیری معکوس: ۲۹

فصل ۳ اطلاع شانون مقادیر و زمان رکوردها در توزیع t - استودنت..... ۴۰

۳-۱- مقدمه ۴۱

۳-۱-۱- توزیع t - استودنت ۴۱

۳-۱-۲- روش تولید داده تصادفی از توزیع t - استودنت ۴۴

۳-۱-۳- تابع توزیع تجمعی t - استودنت ۴۴

۴۴	۳-۱-۴- آنتروپی توزیع t -استودنت
	۳-۲- اطلاع شانون مقادیر و زمان رکوردها توزیع t -استودنت بر اساس یک نمونه n تایی در
۴۸	مدل کلاسیک
۵۱	۳-۳- اطلاع شانون مقادیر و زمان رکوردها توزیع t -استودنت در طرح نمونه گیری معکوس ...
۵۵	۳-۴- نتیجه گیری و پیشنهادات.....
۶۳	منابع و مراجع

فهرست جداول

35	جدول ۱-۲: اطلاع شانون در مدل کلاسیک برای توزیع یکنواخت
۳۵	جدول ۲-۲: اطلاع شانون در طرح نمونه گیری معکوس برای توزیع یکنواخت
۳۹	جدول ۳-۲: اطلاع شانون در مدل کلاسیک برای توزیع نمایی
39	جدول ۴-۲: اطلاع شانون در طرح نمونه گیری معکوس برای توزیع نمایی
	جدول ۱-۳: ($H_M^U(n) = H_M^L(n)$) اطلاع شانون آخرین رکورد بالایی (پایینی) در مدل کلاسیک از
50	توزیع t - استودنت
	جدول ۲-۳: ($H_{RT}^U(n) = H_{RT}^L(n)$) اطلاع شانون مقادیر و زمان رکوردها بالایی (پایینی) در مدل
50	کلاسیک از توزیع t - استودنت
۵۲	جدول ۳-۳: $\psi_f^U(m)$

جدول ۴-۳: $(h_M^U(m) = h_M^L(m))$ اطلاع شانون آخرین رکورد بالایی (پایینی) در طرح نمونه گیری

53..... معکوس از توزیع t - استودنت

جدول ۵-۳: $(h_R^U(m) = h_R^L(m))$ اطلاع شانون مقادیر رکورد بالایی (پایینی) در طرح نمونه گیری

54..... معکوس از توزیع t - استودنت

جدول ۶-۳: $(h_{RT}^U(m) = h_{RT}^L(m))$ اطلاع شانون مقادیر و زمان رکوردها بالایی (پایینی) در طرح

54..... نمونه گیری معکوس از توزیع t - استودنت

فهرست شکل‌ها

شکل ۱-۳: نمودار تابع چگالی احتمال t - استودنت..... ۴۳

فصل ۱

مفاهیم مقدماتی

۱-۱- مقدمه

بشر از سال‌های دور با مفاهیم آماری و احتمالی مأنوس بوده است. گرم‌ترین روز، بلندترین پرش در بسکتبال حرفه‌ای، پایین‌ترین ارقام بازار سهام، احتمال باریدن برف یا باران، احتمال رو آمدن یک سکه در یک یا چند بار پرتاب آن، احتمال برنده شدن در یک شرط بندی و مانند آن، مثال‌هایی از مفاهیم احتمالی هستند که از دیر باز مورد توجه انسان بوده‌اند. معمولاً تمام افراد به مقادیر رکورد علاقه دارند، رکوردها از موضوعاتی هستند که اخیراً در اکثر علوم و فنون مورد مطالعه قرار گرفته‌اند. سؤالاتی از قبیل این که قیمت بورس بازار تهران چه مقدار بالاتر از قیمت امروز خواهد رفت؟ آیا می‌توان بزرگی زلزله‌ای مخرب‌تر از زلزله بم در سال ۱۳۸۲ را به مقیاس ریشتر، پیش‌بینی کرد؟ سیلی مهیب‌تر از سونامی اخیر چه قدرتی دارد؟ همگی، سؤالاتی هستند که ذهن هر انسانی را به سمت خود مشغول خواهند کرد. برای پاسخ به این سؤالات و بررسی جزئیات آن‌ها، بدون شک نیازمند علمی هستیم که بتواند ساختار نظری این پدیده را مورد مطالعه قرار دهد و این موضوع با استفاده از علم آمار و شاخه نظریه توزیع‌ها و گرایش رکوردها قابل بررسی است. به عنوان نمونه‌هایی از کاربرد رکوردها می‌توان به موارد زیر اشاره نمود:

۱- سیل‌های مصیبت‌آفرین و زمین‌لرزه‌های ویران‌گر جزو مواردی هستند که نقش مطالعه رکوردهای بالا را به وضوح برجسته می‌سازند. دانستن زمان وقوع سیلی مخرب‌تر از آخرین سیل ویران‌گر و یا دانستن شدت و زمان رخداد زلزله‌ای مهیب‌تر از آخرین زمین‌لرزه مصیبت‌آفرین، اطلاعاتی است که چه به لحاظ امنیتی و چه به لحاظ ساخت و ساز سدها، ساختمان‌ها و برج‌ها حائز اهمیت می‌باشند که البته همگی این موارد در قالب مطالعه آماره‌های رکوردی، مورد توجه قرار می‌گیرند.

۲- غالباً عموم مردم پیگیر داده‌های رکوردی هستند. مواردی نظیر:

- ❖ بالاترین و پایین‌ترین دما در ماه گذشته
- ❖ بالاترین یا پایین‌ترین آلودگی
- ❖ بیشترین آمار تصادفات در ایام نوروزی
- ❖ مخرب‌ترین زلزله‌ای که تا به حال رخ داده است
- ❖ بالاترین نرخ تورم
- ❖ بالاترین درصد شرکت‌کنندگان در انتخابات
- ❖ بیشترین و کمترین تغییرات در قیمت سهام
- ❖ کمترین میزان ضرر احتمالی در معاملات بورس
- ❖ کمترین میزان بارندگی فصلی
- ❖ کمترین زیان متحمل از انتخاب‌های نادرست

برای بررسی و پیش‌بینی مواردی مانند فوق، نیاز به استفاده از داده‌های رکوردی و به کار بردن روش‌ها و فنون آماری داریم. از دیگر کاربردهای رکوردها می‌توان به کاربرد آن‌ها در رویدادهای ورزشی، بیماری‌های همه‌گیر، تحلیل‌های هواشناسی، آب‌شناسی، زلزله‌شناسی، بحث‌های مربوط به نفت و استخراج معادن، اقتصاد، مطالعات مربوط به فرآیندهای پوآسن، قابلیت اطمینان و کنترل کیفیت، اشاره نمود. در آزمون طول عمر، ممکن است به برآورد هزینه احتمالی دوره ضمانت علاقه‌مند باشیم. به علاوه در مسائل کنترل کیفیت، مقدار مینیمم مشاهده

شده و زمان انجام آزمایش می‌تواند در بهینه‌سازی فرآیند تولید و تعیین حدود تحمل یک چندک جامعه مهم باشد [۱].

کار بر روی مقادیر رکورد نسبتاً پیچیده است اما مشکل دیگر به ماهیت ذاتی رکوردها مربوط می‌شود که کمیاب بودن آنهاست. لذا در نمونه‌های کوچک، توسعه تکنیک‌های استنباطی شامل داده‌های رکوردی مشکل است [۲].

مطالعه مقادیر رکورد به وسیله افراد زیادی انجام شده است. استوارت^۱ [۳] در یافتن کاربردهای آماری رکوردها در استنباط پیشگام بوده است. بارتون^۲، مالوز^۳ و دیوید^۴ [۳] در زمینه کار روی دنباله رکوردها پیشرو بوده‌اند. رنی برخی از اولین قضایای حدی را معرفی کرد. اصطلاح نظریه مجانبی^۵ را می‌توان به تاتا^۶ [۴]، نیوتز^۷ [۳]، گالامبوس^۸ [۵]، رزنیک^۹ [۳]، گلدی^{۱۰} [۳]، شوروک^{۱۱} [۳]، دی‌هان^{۱۲} و دی‌هولز^{۱۳} [۳] منتسب دانست، کار آنها در گستره وسیعی از نشریات و متون علمی پراکنده است و چندین کتاب در مورد آماره‌های ترتیبی و فرآیندهای مرتبط با مقادیر رکورد دارند.

خواص داده‌های رکوردها به طور فراگیر در متون علمی مورد مطالعه قرار گرفته‌اند. چاندلر در سال ۱۹۵۲ برای اولین بار رفتار تصادفی مقادیر رکورد ناشی از «مدل رکورد کلاسیک»^{۱۴} را مطالعه کرد. در این مدل بر اساس یک نمونه تصادفی از یک توزیع پیوسته، تعداد،

¹ Stuart

² Barton

³ Mallows

⁴ F.N. David

⁵ Asymptotic theory

⁶ Tata

⁷ Neuts

⁸ Galambos

⁹ Resnick

¹⁰ Goldie

¹¹ Shorrock

¹² De Haan

¹³ Deheuvels

¹⁴ Classical record model

مقادیر و زمان‌های رکورد مورد مطالعه قرار می‌گیرند. سامانیگو^{۱۵} و ویتاکر^{۱۶} در سال ۱۹۸۶ طرح نمونه‌گیری معکوس را مورد بررسی قرار دادند، که در آن دنباله‌ای از متغیرهای تصادفی (iid) تا زمانی مشاهده می‌شوند که تعداد مقادیر ثابتی از رکوردها به دست آیند. استنباط ناپارامتری بر مبنای رکوردهایی از طرح‌های نمونه‌گیری معکوس به وسیله سامانیگو و ویتاکر [۶]، احمدی^{۱۷} و ارقامی^{۱۸} [۷]، احمدی و بالا کریشنان^{۱۹} در سال ۲۰۰۴ انجام شد. امروزه تعداد بسیار زیادی مقاله و چندین کتاب در مورد داده‌های رکوردی وجود دارند، اما توجه زیادی به خواص آماره‌های رکوردی نشده است. خوانندگان علاقه‌مند می‌توانند به کتاب‌های آرنولد^{۲۰} [۸] و نوزروف^{۲۱} [۹] و مرجع‌های اشاره شده در آن‌ها مراجعه کنند.

احمدی و ارقامی [۱۰]، میزان اطلاع فیشر^{۲۲} (FI) را در داده‌های رکوردی بررسی کردند، هافمن^{۲۳} و ناگارجا^{۲۴} [۱۱] اطلاع فیشر را هم در داده‌های رکورد بالایی و هم پایینی در یک مدل رکورد کلاسیک و نیز برای طرح نمونه‌گیری معکوس^{۲۵} محاسبه کردند. اطلاع فیشر فقط برای توزیع‌هایی معتبر است که شرایط نظم را برآورده کنند. معیارهای دیگری از اطلاع وجود دارند که به این شرایط نیازی ندارند. از بین آن‌ها «آنتروپی شانون^{۲۶}» مشهورترین معیار است. کلمه آنتروپی از کلمه یونانی "entropien" به معنای "مضمون قابل تغییر" یا "محتوای تغییر-پذیر" گرفته شده است. در سال ۱۸۷۰ دو فیزیک‌دان به نامهای گیبز^{۲۷} و بلتزن^{۲۸} معیارهای

¹⁵ Samaniego

¹⁶ Whitaker

¹⁷ Ahmadi

¹⁸ Arghami

¹⁹ Balakrishnan

²⁰ Arnold

²¹ Nevzorov

²² Fisher information

²³ Hafman

²⁴ Nagaraja

²⁵ Inverse sampling plan

²⁶ Shannon entropy

²⁷ J.W. Gibbs

²⁸ L. Boltzmann

آنتروپی را پیشنهاد کردند که به تعمیم‌های زیادی منتهی گردید، که بهترین آنها معیار آرایه شده توسط تسالیس^{۲۹} [۱۲] بود.

اولین مطالعه در مورد معیارهای آنتروپی توسط نایکوئیست^{۳۰} [۱۳، ۱۴] و هارتلی^{۳۱} [۱۵] انجام شد. در سال ۱۹۴۸ کلود شانون، مهندس برق، در حین کار در آزمایشگاه تلفن بل تصمیم گرفت ماهیت آماری «اطلاعات از دست رفته» در سیگنال‌های خط تلفن را به صورت کمی بیان کند. برای انجام این کار شانون مفهوم کلی آنتروپی اطلاع را ارائه داد، که سنگ بنای بنیادی نظریه اطلاع است.

آنتروپی شانون یک متغیر تصادفی X ، یک معیار ریاضی اطلاع است، که متوسط کاهش عدم اطمینان X را اندازه می‌گیرد. به دلیل ماهیت توصیفی آنتروپی شانون، عبارات تحلیلی برای توزیع‌های یک متغیره توسط لازو^{۳۲} و راتی^{۳۳} [۲۸]، کاور^{۳۴} و توماس^{۳۵} [۱۷] بدست آورده شده‌اند. در مقاله هریس^{۳۶} [۱۸]، آنتروپی شانون به عنوان یک کمیت توصیفی از توزیع احتمال متناظر در نظر گرفته شده است. در حالی که کاربردپذیری آن به عنوان یک معیار توصیفی توسط نایاک^{۳۷} [۱۹] توجیه شده است. از زمان آرایه مفهوم آنتروپی توسط شانون، این مفهوم به روش‌های متفاوتی توسط محققان مختلف مانند رنی^{۳۸} (۱۹۶۱) تعمیم یافته و در چندین زمینه و رشته مورد استفاده قرار گرفته است.

آنتروپی در شاخه‌های گوناگون آمار و رشته‌های مرتبط نیز استفاده شده است و امروزه بخش جدانشدنی از آمار و احتمال محسوب می‌شود [۲۰]. برای توزیع‌های چندمتغیری،

²⁹ Tsallis

³⁰ Nyquist

³¹ Hartley

³² Lazo

³³ Rathie

³⁴ Cover

³⁵ Thomas

³⁶ Harris

³⁷ Nayak

³⁸ Renyi

فرمول‌هایی برای آنتروپی شانون در مقالات احمدی و گوخال^{۳۹}[۳]، داربلی^{۴۰} و واجا^{۴۱}[۱۱] ارائه شده‌اند. مطالعه مدت زمان رویدادها، یک موضوع مورد علاقه مشترک برای تجارت، اقتصاد، بیمه، تحلیل بقا و تعداد زیادی از رشته‌هاست. وقتی موضوع مطالعه مدت زمان چیزی غیر از طول عمر باشد (مثلاً زمان تحقیق، دوره بیکاری)، نقطه زمانی فعلی نقش «سن»^{۴۲} را ایفا می‌کند. به طور کلی، معیارهای اطلاع برای هرگونه توزیع پیوسته قابل اعمال هستند. مثلاً برای توزیع‌های مزد، درآمد و منابع طبیعی پایان پذیر مانند نفت،^{۴۳} حداقل دستمزد، خط فقر^{۴۴} و میزان نفت استخراج شده تا به حال، به ترتیب نقش سن فعلی را ایفا می‌کند.

چندین محقق مانند اسدی [۵]، بزونس^{۴۵}[۸]، دیکرسنزو^{۴۶}[۲۱] و لانگوباردی^{۴۷}[۲۲]، و ابراهیمی^{۴۸}[۲۳]، [۲۴] توابع اطلاع را بررسی کرده‌اند که سن را تک‌متغیری در نظر گرفته‌اند. در نظر گرفتن سن به نگرش‌های مهمی در مورد مدل‌های طول عمر منجر شده است مانند مقالات ابراهیمی و کرمانی [۲۵] در مورد تعیین مشخصات اطلاع مخاطره نسبی^{۴۹} و تعیین مشخصات اطلاع ماکزیمم آنتروپی.

در نظریه احتمال و نظریه اطلاع، اطلاع متقابل^{۵۰} دو متغیر تصادفی، کمیتی است که وابستگی متقابل دو متغیر را اندازه‌گیری می‌کند. در اصل معیار اطلاع متقابل، اطلاعی را اندازه می‌گیرد که دو متغیر X و Y در آن سهمیم هستند، در واقع این نوع اطلاع اندازه‌گیری می‌کند که چه میزان دانستن یکی از این متغیرها عدم اطمینان را در مورد متغیر دیگر کاهش می‌دهد.

³⁹ Gokhale

⁴⁰ Darbelly

⁴¹ Vajda

⁴² Age

⁴³ Petroleum

⁴⁴ Minimum Wage

⁴⁵ Bezunce

⁴⁶ Dicrescenzo

⁴⁷ Longobardi

⁴⁸ Ebrahimi

⁴⁹ Proportional hazard

⁵⁰ Mutual information

فصل اول : مفاهیم مقدماتی

به عنوان مثال، اگر X و Y مستقل باشند، آنگاه دانستن X هیچ اطلاعی در مورد Y و بالعکس نمی‌دهد، بنابراین اطلاع متقابل X و Y برابر صفر است. از سوی دیگر اگر X و Y یکسان باشند، آنگاه تمام اطلاع انتقال یافته توسط X با Y مشترک است و دانستن X مقدار Y را مشخص می‌کند و بالعکس. در نتیجه، اطلاع متقابل مشابه عدم قطعیت موجود در Y (یا X) به تنهایی می‌باشد، یعنی اگر X و Y با هم برابر باشند، آنگاه آنتروپی یکسانی دارند. اطلاع متقابل، وابستگی بین توزیع مشترک X و Y را به صورت کمی بیان می‌کند و مشخص می‌کند که توزیع مشترک در صورتی که X و Y مستقل بودند، چگونه می‌شد. همچنین می‌توان نشان داد این اطلاع صفر است اگر و فقط اگر X و Y متغیرهای تصادفی مستقل باشند.

اخیراً چند محقق، براتپور^{۵۱}[۲۶]، زاهدی^{۵۲} و شکیل^{۵۳}[۲۷]، آنتروپی شانون را در داده‌های رکورد بررسی کرده‌اند. این آنتروپی را می‌توان با موفقیت برای کمی سازی میزان اطلاع (در رابطه با خانواده توزیع‌ها) استفاده کرد. در این پایان نامه مفهوم آنتروپی را به داده‌های رکورد تعمیم می‌دهیم و فرمول‌های کلی را برای اطلاع شانون در داده‌های رکورد برای نمونه با اندازه ثابت و طرح نمونه‌گیری معکوس بدست می‌آوریم.

در این قسمت، تعریف مقدار و زمان رکورد را ارایه می‌کنیم و توزیع‌های مربوط به مقادیر رکوردها، برای هر دو مدل کلاسیک و طرح نمونه‌گیری معکوس که توسط هافمن و ناگارجا ارایه شده‌اند، مورد بررسی قرار می‌گیرند [۱۶].

فرض کنید $\{X_n : n = 1, 2, \dots\}$ یک دنباله مستقل و هم‌توزیع (iid) از متغیرهای تصادفی با تابع توزیع تجمعی کاملاً پیوسته $F(x)$ و تابع چگالی احتمال $f(x)$ باشد. متغیر X_j را یک مقدار رکورد بالا می‌نامیم، اگر مقدار آن بیشتر از همه مشاهدات قبلی باشد. بنابراین X_j یک

⁵¹ Baratpour

⁵² Zahedi

⁵³ Shakil

فصل اول : مفاهیم مقدماتی

رکورد بالاست، اگر $X_j > X_i$ برای هر $i < j$ باشد. برای راحتی فرض می‌کنیم که X_j در زمان j ام مشاهده شده است، در این صورت دنباله زمان رکوردهای بالایی^{۵۴} $\{T_n, n \geq 1\}$ به صورت زیر تعریف می‌شود:

$$P(T_1 = 1) = 1,$$

و برای $n > 1$ داریم:

$$T_n = \min\{j > T_{n-1}\}.$$

بنابراین دنباله مقادیر رکورد بالایی به صورت زیر تعریف می‌شود:

$$R_n = X_{T_n}, \quad n = 1, 2, 3, \dots$$

همچنین تعداد رکوردهای بالایی در میان $X_1, X_2, \dots, X_n$ را با N_n نشان می‌دهیم که یک متغیر تصادفی می‌باشد. به طور مشابه می‌توان مقادیر، زمان و تعداد رکوردهای پایینی را در دنباله $\{X_n\}$ تعریف کرد.

۱-۱-۱- مدل کلاسیک^{۵۵}

فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی با تابع توزیع تجمعی $F(x)$ و تابع چگالی احتمال $f(x)$ باشد، در این صورت توزیع توأم رکوردهای بالایی $R_1, R_2, \dots, R_{N_n}$ ، زمانهای رکورد $T_1, T_2, \dots, T_{N_n}$ و تعداد رکوردها N_n به صورت زیر مشخص می‌شود [۲].

$$f(r_1, \dots, r_m, t_1, \dots, t_m, N_n = m) = \prod_{i=1}^m f(r_i) \prod_{i=1}^m F^{\delta_i}(r_i), \quad (1-1)$$

که در آن $r_1 < \dots < r_m$ ، $t_1 = 1 < \dots < t_m$ ، $\delta_i = t_{i+1} - t_i - 1$ و $t_{m+1} = n + 1$.

⁵⁴ Uper record time sequence

⁵⁵ Classical model

بدیهی است که $\sum_{i=1}^m \delta_i = n - m$ و برای هر i ، $\delta_i \geq 0$.

با جمع بستن روی زمان‌های رکورد می‌توان توزیع مقادیر و تعداد رکوردهای بالایی را به صورت زیر بدست آورد

$$f(r_1, \dots, r_m, N_n = m) = \prod_{i=1}^m f(r_i) \sum_{(\delta_1, \dots, \delta_m)} \prod_{i=1}^m F^{\delta_i}(r_i), \quad (2-1)$$

برای $r_1 < \dots < r_m$ ، $m = 1, \dots, n$ و $\sum_{i=1}^m \delta_i = n - m$ و برای هر i ، $\delta_i \geq 0$. علاوه بر این آخرین

رکورد همان آماره ترتیبی n ام در نمونه می‌باشد بنابراین

$$f_n(x) = n F^{n-1}(x) f(x). \quad (3-1)$$

۱-۲-۱- طرح نمونه‌گیری معکوس

در مدل طرح نمونه‌گیری معکوس، در ابتدا تعداد رکوردها، m ، را مشخص می‌کنیم و نمونه‌گیری را تا رسیدن به m رکورد ادامه می‌دهیم. فرض کنید $X_1, X_2, X_3, \dots$ یک دنباله از متغیرهای تصادفی هم‌توزیع و مستقل با تابع توزیع پیوسته $F(x)$ و تابع چگالی احتمال $f(x)$ برای مقدار ثابت m باشند. احتمال مشترک^{۵۶} [۱۶] مقادیر رکورد بالایی^{۵۷} $R_1, R_2, \dots, R_m$ و زمان‌های رکورد $T_1, T_2, \dots, T_{N_n}$ به صورت زیر است:

$$f(r_1, \dots, r_m, t_1, \dots, t_m) = \prod_{i=1}^m f(r_i) \prod_{i=1}^{m-1} F^{\delta_i}(r_i), \quad (4-1)$$

که در آن، $r_1 < \dots < r_m$ ، $t_1 = 1 < \dots < t_m$ و $\delta_i = t_{i+1} - t_i - 1$.

از آنجایی که δ_i ها کراندار نیستند، چگالی مشترک مقادیر رکوردهای $R_1, R_2, \dots, R_m$ به صورت

⁵⁶ Joint probability

⁵⁷ Uper record data

$$f(r_1, \dots, r_m) = \prod_{i=1}^{m-1} \frac{f(r_i)}{1-F(r_i)} f(r_m), \quad r_1 < \dots < r_m, \quad (5-1)$$

است. همچنین تابع چگالی احتمال حاشیه‌ای آخرین رکورد،⁵⁸ عبارت است از

$$f(r_m) = \frac{(-\log(1-F(r_m)))^{m-1}}{(m-1)!} f(r_m). \quad (6-1)$$

۱-۲- آنروپی

پیشامد E را در نظر می‌گیریم که می‌تواند اتفاق بیفتد. چقدر تعجب می‌کنیم اگر بشنویم که E اتفاق افتاده است؟ به نظر منطقی است که تصور کنیم میزان تعجب حاصل از اطلاع وقوع E بایستی به احتمال وقوع آن بستگی داشته باشد. به عنوان مثال، اگر آزمایش شامل پرتاب یک جفت تاس باشد، آنگاه اگر بشنویم که جمع اعداد ظاهر شده زوج است زیاد متعجب نمی‌شویم، زیرا احتمال وقوع آن برابر با $\frac{1}{2}$ است، در حالی که اگر بشنویم جمع اعداد ظاهر شده برابر با 12 است بیشتر متعجب می‌شویم، چون احتمال وقوع آن برابر با $\frac{1}{36}$ است. اگر میزان تعجب را کمی کنیم و روی این اصل توافق کنیم که میزان تعجب از وقوع پیشامد E فقط به احتمال وقوع آن بستگی دارد، آنگاه می‌توان آن را با نماد $S(p)$ نشان داد، یعنی میزان تعجبی که از وقوع پیشامدی با احتمال p حاصل می‌گردد. حال فرض می‌کنیم که $S(p)$ برای همه مقادیر $0 < p \leq 1$ تعریف شده باشد. در این جا به چهار اصل که در مبحث آنروپی مهم می‌باشند، اشاره می‌کنیم. وقتی که بشنویم پیشامد حتمی که باید اتفاق می‌افتاده رخ داده است، تعجب ما از وقوع آن صفر است، یعنی

$$S(1) = 0 \quad ۱.$$

⁵⁸ Marginal probability density function

فصل اول : مفاهیم مقدماتی

۲. $S(p)$ تابعی اکیداً نزولی از p است یعنی اگر $p < q$ ، آنگاه

$$S(p) > S(q).$$

۳. $S(p)$ تابعی پیوسته از p است.

دو پیشامد مستقل E و F را در نظر می‌گیریم به طوری که $P(E) = p$ و $P(F) = q$ بنابراین $P(EF) = pq$. میزان تعجب حاصل از اطلاع وقوع E و F با $S(pq)$ برابر است، چون E و F از یکدیگر مستقلند و اطلاع از وقوع یا عدم وقوع پیشامد E تأثیری در احتمال وقوع پیشامد F ندارد. بنابراین

۴. برای $0 < p \leq 1$ و $0 < q \leq 1$

$$S(pq) = S(p) + S(q).$$

قضیه ۱-۲-۱ اگر $S(\cdot)$ در اصول ۱ تا ۴ صدق کند، آنگاه

$$S(p) = -C \log_2 p,$$

به طوری که C یک عدد دلخواه مثبت با مقدار $\ln 2$ است. بنابراین $S(p) = -\ln p$. حال متغیر تصادفی X که یکی از مقادیر $x_1, \dots, x_n$ را با احتمال‌های $p_1, \dots, p_n$ انتخاب می‌کند، در نظر می‌گیریم. چون $-\ln p_i$ نشان‌دهنده میزان تعجب حاصل از این است که X مقدار x_i را می‌گیرد، در نتیجه میزان تعجب از اطلاع در مورد مقدار متغیر تصادفی X برابر است با

$$H(X) = -\sum_{i=1}^n p_i \ln p_i. \quad (7-1)$$

واضح است که

$$H(X) = E(-\ln f(X)), \quad (8-1)$$

فصل اول : مفاهیم مقدماتی

که در آن، $f(x)$ تابع جرم احتمال متغیر X می باشد.

کمیت $H(X)$ در نظریه اطلاع، به عنوان آنتروپی متغیر تصادفی X شناخته می شود. در حالتی که یکی از مقادیر $p_i = 0$ باشد، آن گاه $0 \ln 0 = 0$ در نظر می گیریم. می توان نشان داد که $H(X)$ زمانی حد اکثر می شود که همه p_i ها برابر باشند [۲].

از آنجایی که $H(X)$ نشان دهنده متوسط میزان تعجبی است که در مورد اطلاع از مقدار X حاصل می گردد، می توان آن را به عنوان نشان دهنده میزان عدم قطعیتی که برای مقدار X وجود دارد نیز تفسیر نمود. در واقع در نظریه اطلاع $H(X)$ را متوسط اطلاعات به دست آمده برای مقدار مشاهده شده X می نامند. در مجموع، متوسط میزان تعجب حاصل شده از X ، عدم قطعیت X یا متوسط میزان اطلاعات حاصل از مقدار مشاهده شده X ، همگی مفاهیم یکسانی هستند که از دیدگاه های متفاوتی بیان می شوند.

حال دو متغیر تصادفی X و Y را که به ترتیب مقادیر $x_1, \dots, x_n$ و $y_1, \dots, y_m$ می گیرند، با تابع چگالی احتمال توأم^{۵۹} زیر در نظر می گیریم:

$$f(x_i, y_j) = p\{X = x_i, Y = y_j\}, \quad \begin{matrix} i = 1, \dots, n, \\ j = 1, \dots, m, \end{matrix}$$

آن گاه میزان عدم قطعیت^{۶۰} برای مقدار بردار تصادفی^{۶۱} (x, y) که با $H(X, Y)$ نشان داده می شود، برابر است با

$$H(X, Y) = - \sum_i \sum_j f(x_i, y_j) \ln f(x_i, y_j). \quad (9-1)$$

⁵⁹ Joint mass function

⁶⁰ Uncertainty

⁶¹ Random vector

فصل اول : مفاهیم مقدماتی

اگر مقدار $Y = y_j$ مشاهده شده باشد، در این حالت مقدار عدم قطعیت باقیمانده X برابر است با

$$H(X | y_j) = - \sum_i f(x_i | y_j) \ln f(x_i | y_j), \quad (10-1)$$

به طوری که

$$f(x_i | y_j) = p\{X = x_i | Y = y_j\}.$$

بنابراین متوسط مقدار عدم قطعیت که بعد از مشاهده Y برای X باقی می ماند برابر است با

$$\begin{aligned} H(X | Y) &= \sum_j H(X | y_j) f_Y(y_j) \\ &= - \sum_i \sum_j f(x_i, y_j) \ln f(x_i | y_j), \end{aligned} \quad (11-1)$$

به طوریکه

$$f_Y(y_j) = p\{Y = y_j\}$$

روابط بالا ارتباط بین $H(X, Y)$ را با $H(Y)$ و $H(X | Y)$ نشان می دهد. این روابط بیان می دارند که عدم قطعیت برای مقادیر X و Y برابر است با مجموع عدم قطعیت Y و متوسط عدم قطعیتی که بعد از مشاهده Y برای X باقی می ماند:

$$H(X, Y) = H(Y) + H(X | Y).$$

یکی از نتایج اساسی در نظریه اطلاع این است که میزان عدم قطعیت در متغیر تصادفی X ، به طور متوسط، با مشاهده دومین متغیر تصادفی Y کاهش می یابد. در نتیجه

$$H(X, Y) \leq H(X) + H(Y).$$

و نامساوی به تساوی تبدیل می شود اگر و تنها اگر X و Y از هم مستقل باشند.

فصل اول : مفاهیم مقدماتی

اطلاع مشترک $H(X,Y)$ زمانی به بیشترین مقدار خواهد رسید که X و Y از هم مستقل باشند. بنابراین افزایش وابستگی بین X و Y ، اطلاع مشترک را کاهش می‌دهد.

تعریف ۱-۲-۲ اگر X یک متغیر تصادفی با تابع توزیع پیوسته $F(X)$ و تابع چگالی پیوسته $f(x)$ باشد، آنترپی شانون به صورت زیر تعریف می‌شود:

$$H(X) = - \int_{-\infty}^{+\infty} f(x) \ln f(x) dx. \quad (1-12)$$

تذکر باید توجه داشته باشیم که در مورد متغیر تصادفی X ، $H(X)$ تحت یک تبدیل یک به یک از X همیشه نامنفی و یکسان است، اما هنگامی که X یک متغیر تصادفی پیوسته است، آنترپی تحت تبدیل یکسان، یک به یک نیست و بین $(-\infty, +\infty)$ مقدار می‌گیرد.

لم ۱-۲-۳ فرض کنید Z یک متغیر تصادفی با تابع توزیع F_Z ، تابع چگالی f_Z و اطلاع شانون، $H(Z)$ باشد. در این صورت

❖ اگر برای هر $\theta \in \mathbb{R}$ ، $X = Z + \theta$ ، آن‌گاه $F_X(x) = F_Z(x - \theta)$ ، $f_X(x) = f_Z(x - \theta)$

$$\text{و } H(X) = H(Z).$$

❖ اگر برای هر $\theta > 0$ ، $X = \theta Z$ ، آن‌گاه $F_X(x) = F_Z(\frac{x}{\theta})$ ، $f_X(x) = \frac{1}{\theta} f_Z(\frac{x}{\theta})$ و

$$H(X) = \ln \theta + H(Z)$$

❖ اگر برای هر $\theta > 0$ ، $X = Z^\theta$ ، آن‌گاه $F_X(x) = F_Z(x^{\frac{1}{\theta}})$ ، $f_X(x) = \frac{1}{\theta} x^{\frac{1}{\theta}-1} f_Z(x^{\frac{1}{\theta}})$ و

$$H(X) = \ln \theta + H(Z) - (\frac{1}{\theta} - 1) E(\ln X)$$

فصل اول : مفاهیم مقدماتی

اگر (X, Y) یک بردار تصادفی با تابع چگالی احتمال توأم $f(x, y)$ باشد، اطلاع مشترک شانون^{۶۲} به صورت زیر تعریف می‌شود

$$H(X, Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \ln f(x, y) dx dy. \quad (۱۳-۱)$$

بنابراین اطلاع شرطی^{۶۳} $X | Y$ به صورت زیر به دست می‌آید

$$H(X | Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \ln f(x | y) dx dy. \quad (۱۴-۱)$$

باید به این نکته توجه داشته باشیم، اگر چه برای متغیرهای تصادفی گسسته $H(X, Y) > 0$ می‌باشد، اما برای متغیرهای پیوسته $H(X, Y)$ ممکن است منفی باشد. با این حال نتایج بالا هنوز هم برقرار هستند.

تعریف ۱-۲-۴ اگر X و Y هر دو متغیر تصادفی گسسته باشند، اطلاع دوطرفه بین X و Y به صورت زیر تعریف می‌شود:

$$I(X, Y) = H(X) - H(X | Y). \quad (۱۵-۱)$$

و اگر X و Y هر دو پیوسته باشند، داریم:

$$I(X, Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \ln \frac{f(x, y)}{f(x)f(y)} dx dy. \quad (۱۶-۱)$$

در زیر برخی از خواص $I(X, Y)$ را فهرست می‌کنیم [۲]:

- خاصیت نامنفی بودن^{۶۴} : $I(X, Y) \geq 0$.

⁶² Joint shannon information

⁶³ Conditional information

⁶⁴ Non-negativity

- خاصیت تغییرناپذیری^{۶۵}: اگر $V = h(Y)$ و $U = g(X)$ که در آن h و g هر دو توابعی

یک به یک هستند آن گاه $I(X, Y) = I(U, V)$.

- $I(X, Y) = 0$ اگر و تنها اگر X و Y مستقل باشند.

- $I(X, Z) \leq I(X, Y)$ اگر X مستقل از Z و Y باشد.

- $I(X, Y) = H(X) + H(Y) - H(X, Y)$

چندین تعمیم از تعریف آنترپی شانون پیشنهاد شده اند، که دو مورد از آنها عبارتند از:

(۱) رنی [۲۹] آنترپی شانون را به صورت زیر تعمیم داد:

$$H(X, \beta) = \frac{1}{1-\beta} \ln \left(\int_{-\infty}^{+\infty} f^{\beta}(x) dx \right), \quad \beta \neq 1, \beta > 0. \quad (۱۷-۱)$$

توجه داشته باشید که اگر β به سمت ۱ میل کند اطلاع رنی به اطلاع شانون نزدیک می شود.

علاوه بر این، $H(X, \beta)$ به طور یکنواخت یک تابع کاهشی از β است [۷].

(۲) تسالیس [۱۲] آنترپی را به صورت زیر تعریف کرد:

$$\tilde{H}(X, \beta) = \frac{1}{\beta-1} \left(1 - \int_{-\infty}^{+\infty} f^{\beta}(x) dx \right), \quad \beta \neq 1, \beta > 0. \quad (۱۸-۱)$$

واضح است که

$$\tilde{H}(X, \beta) = \frac{1}{\beta-1} (1 - e^{(1-\beta)H(X, \beta)}).$$

⁶⁵ Invariance

۱-۳- معایب و مزایای استفاده از اطلاع شانون به جای عدم قطعیت

از آنجایی که عبارت اطلاع شانون در فیزیک و مهندسی بیش از ۱۵۰ سال استفاده شده است، تعدادی از منتقدان اعتقاد دارند که بحث روی ماکزیمم کردن اطلاع شانون باید تنها متعلق به فیزیک و مهندسی باشد برخی علاقه به بهینه سازی اطلاع شانون ندارند و بیشتر به قسمت مفهومی بودن مساله توجه دارند به همین دلیل بیشتر از عدم قطعیت استفاده می کنند. و مزایای استفاده از عدم قطعیت به جای اطلاع شانون این است که عدم قطعیت با همه انشعاباتش یک مفهوم خیلی وسیع دارد و به عنوان یک اندازه، تابعی از $(P_1, P_2, \dots, P_n)$ است. با استفاده از عدم قطعیت، ممکن است به اشتباه تصور کنیم که با همه نوع عدم قطعیت کار می کنیم. بنابراین بهتر است از عبارتی استفاده کنیم که در زندگی روزمره استفاده نمی شود. از این رو، اطلاع شانون به طور نسبی دقیق تر تعریف می شود. اعمال ریاضی را روی اطلاع شانون می توان اجرا کرد، اما عدم قطعیت، یک عبارت کیفی است و نمی توانیم اعمال ریاضی را روی آن اجرا کنیم. به عنوان مثال، استفاده از ماکزیمم عدم قطعیت نسبت به ماکزیمم اطلاع شانون خیلی مشکل تر است.

برای درک بهتر اطلاع شانون انواع آن را به کار می برند. هنگامی که اطلاعاتی درباره میانگین انرژی یا سرعت سیستم داریم، اطلاع شانون را اطلاع شانون ترمودینامیک می نامند. هنگامی که اطلاعات، متغیرهای اقتصادی مثل هزینه ها، قیمت و یا اجاره باشد، اطلاع شانون را اطلاع شانون اقتصادی می نامند. بنابراین، عبارت اطلاع شانون می تواند موضوعات زیادی را نسبت به عبارت عدم قطعیت پوشش دهد.

فصل ۲

اطلاع در داده های رکوردی

فصل دوم: اطلاع در داده های رکوردی

در این فصل که پیش رو داریم به ارائه نتایجی از اطلاع شانون که شامل مقادیر رکورد بالا(پایین) و ارتباط زمان رخداد رکوردها در یک دنباله از متغیرهای تصادفی مستقل پیوسته(مدل کلاسیک) می پردازیم. و سپس اطلاع مقادیر رکورد در طرح نمونه گیری معکوس را بررسی می کنیم. و در پایان این فصل اطلاع شانون را برای چند توزیع معروف محاسبه می کنیم.

۲-۱- اطلاع شانون در مقادیر رکورد.

۲-۱-۱- اطلاع در یک نمونه محدود از مقادیر رکوردها

برای محاسبه مقادیر ماکزیمم رکورد (رکوردهای بالایی) برای نمونه هایی در اندازه ثابت n ، بایستی موارد زیر را محاسبه کنیم:

• $H_{RT}^U(n)$: اطلاع شانون رکوردهای بالایی و زمان های رکورد.

• $H_R^U(n)$: اطلاع شانون رکوردهای بالایی.

• $H_M^U(n)$: اطلاع شانون ماکزیمم رکورد (آخرین رکورد).

بنابراین، به طور مشابه، اطلاع شانون مربوط به رکوردهای پایین وابسته به این سه مورد را با نمادهای $H_{RT}^L(n)$ ، $H_R^L(n)$ و $H_M^L(n)$ ، که به ترتیب به نشان دهنده اطلاع شانون رکوردهای پایینی و زمان های رکورد، اطلاع شانون رکوردهای پایینی و اطلاع شانون مینیمم رکورد هستند، نشان می دهیم.

در یک مدل کلاسیک از آنجایی که تعداد مشاهدات مقدار ثابت n می باشد، بنابراین با استفاده از (۱-۳)، آخرین رکورد (ماکزیمم رکورد) همان ماکزیمم نمونه با تابع چگالی

$$f_n(x) = n F^{n-1}(x) f(x),$$

می باشد، بنابراین می توانیم نشان دهیم

$$\begin{aligned}
 H_M^U(n) &= -\int_{-\infty}^{+\infty} f_n(x) \ln f_n(x) dx \\
 &= -\ln n + \int_{-\infty}^{+\infty} (-\ln f(x) - (n-1) \ln F(x)) n F^{n-1}(x) f(x) dx \\
 &= -\ln n + \frac{n-1}{n} - n \phi_f^U(n-1),
 \end{aligned} \tag{۱-۲}$$

که در آن، $\phi_f^U(i)$ به صورت زیر تعریف می شود:

$$\phi_f^U(i) = \int_{-\infty}^{+\infty} (\ln f(x)) F^i(x) f(x) dx. \tag{۲-۲}$$

همچنین برای مقادیر پایین رکورد داریم

$$H_M^L(n) = -\ln n + \frac{n-1}{n} - n \phi_f^L(n-1), \tag{۳-۲}$$

که در آن، $\phi_f^L(i)$ با جایگذاری $F(x)$ با $1-F(x)$ در $\phi_f^U(i)$ به دست می آید [۳۰].

قضیه ۱-۱-۲ * (مددی و تاتا ۲۰۱۱) اطلاع شانون همه مقادیر رکوردهای بالایی و زمانهای

رکورد برای یک نمونه تصادفی به حجم n (مدل کلاسیک)، به صورت

$$H_{RT}^U(n) = \sum_{i=0}^{n-1} \int_{-\infty}^{+\infty} (-\ln f(x) - i \ln F(x)) F^i(x) f(x) dx \tag{۴-۲}$$

به دست می آید. به طور مشابه، اطلاع شانون متناظر با رکوردهای پایینی و زمان رکوردها

$(H_{RT}^L(n))$ با جایگذاری $1-F(x)$ با $F(x)$ در (۴-۲) به دست می آید [۳۰].

برهان. با مراجعه به تابع چگالی توأم در (۱-۱)، داریم

$$H_{RT}^U(n) = E \left(-\ln f(R_1, \dots, R_{N_n}, T_1, \dots, T_{N_n}, N_n) \right),$$

که در آن، امید ریاضی نسبت به تابع چگالی توأم (۱-۱) می باشد. لذا

$$\begin{aligned}
 H_{RT}^U(n) &= \sum_{m=1}^n E \left[-\ln \left(\prod_{i=1}^m f(R_i) \prod_{i=1}^m F^{\delta_i}(R_i) \right) \right] \\
 &= \sum_{m=1}^n E \left(-\sum_{i=1}^m \ln f(R_i) \right) + \sum_{m=1}^n E \left(-\sum_{i=1}^m \delta_i \ln F(R_i) \right) \\
 &= \sum_{m=1}^n A_1(m) + \sum_{m=1}^n A_2(m) \\
 &= A_1 + A_2.
 \end{aligned} \tag{۵-۲}$$

که A_1 و A_2 به صورت زیر نشان داده می شوند

$$\begin{aligned}
 A_1 &= \sum_{m=1}^n E \left(-\sum_{i=1}^m \ln f(R_i) \right) \\
 A_2 &= \sum_{m=1}^n E \left(-\sum_{i=1}^m \delta_i \ln F(R_i) \right)
 \end{aligned}$$

برای این که بتوانیم رابطه (۴-۲) در قضیه ۱-۱-۲ را نشان دهیم باید A_1 و A_2 را بدست آوریم.

بنابراین، ابتدا مقدار A_1 را محاسبه می کنیم.

$$\begin{aligned}
 A_1(m) &= E \left(-\sum_{i=1}^m \ln f(R_i) \right) \\
 &= \int_{r_1 < \dots < r_m} \left\{ \sum_{(\delta_1, \dots, \delta_m)} \left(-\sum_{i=1}^m \ln f(r_i) \right) \prod_{i=1}^m f(r_i) \times \prod_{i=1}^m F^{\delta_i}(r_i) \right\} dr_1 \dots dr_m
 \end{aligned}$$

بنابراین،

$$A_1(m) = \int_{r_1 < \dots < r_m} \left\{ \left(-\sum_{i=1}^m \ln f(r_i) \right) \prod_{i=1}^m f(r_i) \times \left(\sum_{(\delta_1, \dots, \delta_m)} \prod_{i=1}^m F^{\delta_i}(r_i) \right) \right\} dr_1 \dots dr_m.$$

فصل دوم: اطلاع در داده های رکوردی

توجه داشته باشید $\sum_{i=1}^m \delta_i = n - m$ با شرط $\sum_{i=1}^m \prod_{(\delta_1, \dots, \delta_m)} F^{\delta_i}(r_i)$ تابعی است و برابر است با ضریب

s^{n-m} در بسط سری تیلور $\prod_{j=1}^m \frac{1}{1-F(r_j)s}$ [۲]. همچنین ضریب s^k را در بسط سری تیلور

$g(s)$ با $\langle g(s) \rangle^{(k)}$ نشان می‌دهیم. پس

$$\begin{aligned} A_1(m) &= \int_{r_1 < \dots < r_m} \left\{ \left(-\sum_{i=1}^m \ln f(r_i) \right) \prod_{i=1}^m f(r_i) \times \left\langle \prod_{j=1}^m \frac{1}{1-F(r_j)s} \right\rangle^{(n-m)} \right\} dr_1 \dots dr_m \\ &= \left\langle \int_{r_1 < \dots < r_m} \left(-\sum_{i=1}^m \ln f(r_i) \right) \prod_{j=1}^m \frac{f(r_j)}{1-F(r_j)s} dr_1 \dots dr_m \right\rangle^{(n-m)}. \end{aligned}$$

توجه داشته باشید که تابع زیر انتگرال، نسبت به r_i ها متقارن است. از این رو،

$$\begin{aligned} A_1(m) &= \left\langle \int_{-\infty}^{+\infty} \int_{-\infty}^{r_m} \dots \int_{-\infty}^{r_2} \left(-\sum_{i=1}^m \ln f(r_i) \right) \times \prod_{j=1}^m \frac{f(r_j)}{1-F(r_j)s} dr_1 \dots dr_m \right\rangle^{(n-m)} \\ &= \left\langle \frac{1}{m!} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left(-\sum_{i=1}^m \ln f(r_i) \right) \times \prod_{j=1}^m \frac{f(r_j)}{1-F(r_j)s} dr_1 \dots dr_m \right\rangle^{(n-m)} \\ &= \left\langle \frac{1}{m!} \sum_{i=1}^m \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} (-\ln f(r_i)) \times \prod_{j=1}^m \frac{f(r_j)}{1-F(r_j)s} dr_1 \dots dr_m \right\rangle^{(n-m)} \\ &= \left\langle \frac{1}{m!} \left[\sum_{i=1}^m \int_{-\infty}^{+\infty} (-\ln f(r_i)) \frac{f(r_i)}{1-F(r_i)s} dr_i \right] \times \prod_{j=1, j \neq i}^m \left(\int_{-\infty}^{+\infty} \frac{f(r_j)}{1-F(r_j)s} dr_j \right) \right\rangle^{(n-m)} \end{aligned}$$

که به دلیل تقارن، می‌توانیم $A_1(m)$ را به صورت زیر نشان دهیم:

$$A_1(m) = \left\langle \frac{1}{(m-1)!} \left(\int_{-\infty}^{+\infty} (-\ln f(x)) \frac{f(x)}{1-F(x)s} dx \right) \times \left(\int_{-\infty}^{+\infty} \frac{f(x)}{1-F(x)s} dx \right)^{m-1} \right\rangle^{n-m}.$$

با استفاده از بسط $\frac{1}{1-F(x)s}$ در سری های توانی، داریم

$$A_1(m) = \left\langle \frac{1}{(m-1)!} \left[\int_{-\infty}^{+\infty} (-\ln f(x)) f(x) \left(\sum_{i=0}^{+\infty} F^i(x) s^i \right) dx \right] \times \left(-\frac{1}{s} \log(1-s) \right)^{m-1} \right\rangle^{n-m}$$

$$= \left\langle \sum_{i=0}^{+\infty} \left(s^i \int_{-\infty}^{+\infty} (-\ln f(x)) F^i(x) f(x) dx \right) \times \frac{1}{(m-1)!} \left(-\frac{1}{s} \log(1-s) \right)^{m-1} \right\rangle^{n-m}.$$

از آنجایی که ضریب s^{n-m} با مجموع حاصل از ضرایب s^j در پرانتز اول و ضرایب s^{n-m-j} در

پرانتز دوم برابر است، بنابراین

$$A_1(m) = \sum_{j=0}^{n-m} \left\{ \left\langle \sum_{i=0}^{+\infty} \left(\int_{-\infty}^{+\infty} (-\ln f(x)) F^i(x) f(x) dx \right) s^i \right\rangle^{(j)} \times \left\langle \frac{1}{(m-1)!} \left(-\frac{1}{s} \log(1-s) \right)^{m-1} \right\rangle^{n-m-j} \right\}$$

$$= \sum_{j=0}^{n-m} \left\{ \left\langle \int_{-\infty}^{+\infty} (-\ln f(x)) F^j(x) f(x) dx \right\rangle \times \frac{1}{(m-1)!} \left\langle \frac{(-\log(1-s))^{m-1}}{s^{m-1}} \right\rangle^{n-m-j} \right\}$$

$$= \sum_{i=0}^{n-m} \left\{ \left\langle \int_{-\infty}^{+\infty} (-\ln f(x)) F^i(x) f(x) dx \right\rangle \times \left\langle \frac{(-\log(1-s))^{m-1}}{(m-1)!} \right\rangle^{n-i-1} \right\}.$$

از این رو با گرفتن مجموع بالا روی $A_1(m)$ داریم

$$A_1 = \sum_{m=1}^n A_1(m)$$

$$= \sum_{m=1}^n \sum_{i=0}^{n-m} \left\{ \left(\int_{-\infty}^{+\infty} (-\ln f(x)) F^i(x) f(x) dx \right) \times \left\langle \frac{(-\log(1-s))^{m-1}}{(m-1)!} \right\rangle^{(n-i-1)} \right\}.$$

با تغییر دادن ترتیب مجموعهای بالا، A_1 را می توان به صورت زیر نشان داد:

$$A_1 = \sum_{i=0}^{n-1} \left[\left(\int_{-\infty}^{+\infty} (-\ln f(x)) F^i(x) f(x) dx \right) \times \sum_{m=1}^{n-i} \left\langle \frac{(-\log(1-s))^{m-1}}{(m-1)!} \right\rangle^{(n-i-1)} \right]$$

مجموع دوم ذکر شده در A_1 ، ضریبی از s^{n-i-1} در $\sum_{m=1}^{\infty} \frac{(-\log(1-s))^{m-1}}{(m-1)!}$ می باشد. با استفاده از

بسط سری تیلور عبارت $(-\log(1-s))^k$ را می توان به صورت

$$(-\log(1-s))^k = \left(s + \frac{s^2}{2} + \dots\right)^k,$$

نوشت، که در آن ضریب s^{n-i-1} برای $k > n-i-1$ ، برابر با صفر است. این مجموع در حقیقت برابر است با ضریبی از s^{n-i-1} در

$$\sum_{m=1}^{+\infty} \frac{(-\log(1-s))^{m-1}}{(m-1)!} = \sum_{j=0}^{+\infty} \frac{(-\log(1-s))^j}{(j)!} = e^{-\ln(1-s)} = \frac{1}{1-s} = \sum_{j=0}^{+\infty} s^j,$$

و برابر است با ۱.

بنابراین عبارت اول در (۵-۲)، برابر است با

$$A_1 = \sum_{i=0}^{n-1} \int_{-\infty}^{+\infty} (-\ln f(x)) F^i(x) f(x) dx. \quad (۶-۲)$$

$A_2 = \sum_{m=1}^n A_2(m)$ را می توان به صورت زیر بیان کرد:

$$\begin{aligned} A_2(m) &= E \left(-\sum_{i=1}^m \delta_i \ln F(R_i) \right) \\ &= \int_{r_1 < \dots < r_m} \left\{ \sum_{(\delta_1, \dots, \delta_m)} \left(-\sum_{i=1}^m \delta_i \ln F(r_i) \right) \prod_{j=1}^m f(r_j) \times \prod_{k=1}^m F^{\delta_k}(r_k) \right\} dr_1 \dots dr_m \\ &= \int_{r_1 < \dots < r_m} \left\{ \left(-\sum_{i=1}^m \ln F(r_i) \right) \prod_{j=1}^m f(r_j) \times \left(\sum_{(\delta_1, \dots, \delta_m)} \delta_i \prod_{k=1}^m F^{\delta_k}(r_k) \right) \right\} dr_1 \dots dr_m, \quad (۷-۲) \end{aligned}$$

که در آن $\sum_{i=1}^m \delta_i = n-m$ باید توجه داشته باشید

$$\begin{aligned}
 \sum_{(\delta_1, \dots, \delta_m)} \delta_i \prod_{k=1}^m F^{\delta_k}(r_k) &= \sum_{(\delta_1, \dots, \delta_m)} \frac{F(r_i)}{f(r_i)} \frac{\partial}{\partial r_i} \left(\prod_{k=1}^m F^{\delta_k}(r_k) \right) \\
 &= \frac{F(r_i)}{f(r_i)} \frac{\partial}{\partial r_i} \left\langle \prod_{k=1}^m \frac{1}{1-F(r_k)s} \right\rangle^{(n-m)} \\
 &= \left\langle \frac{F(r_i)s}{(1-F(r_i)s)^2} \prod_{k=1, k \neq i}^m \frac{1}{1-F(r_k)s} \right\rangle^{(n-m)}.
 \end{aligned}$$

حال با استفاده از رابطه های داخل (۷-۲) و استدلال های مشابه قبل، نتیجه می گیریم

$$\begin{aligned}
 A_2(m) &= \left\langle \int_{r_1 < \dots < r_m} \left[\sum_{i=1}^m (-\ln F(r_i)) \frac{f(r_i)F(r_i)s}{(1-F(r_i)s)^2} \times \prod_{j=1, j \neq i}^m \frac{f(r_j)}{1-F(r_j)s} \right] dr_1 \dots dr_m \right\rangle^{(n-m)} \\
 &= \left\langle \frac{1}{(m-1)!} \left(\int_{-\infty}^{+\infty} (-\ln F(x)) \frac{f(x)F(x)s}{(1-F(x)s)^2} dx \right) \times \left(\int_{-\infty}^{+\infty} \frac{f(x)}{1-F(x)s} dx \right)^{m-1} \right\rangle^{(n-m)} \\
 &= \left\langle \frac{1}{(m-1)!} \left(\int_{-\infty}^{+\infty} (-\ln F(x)) \frac{f(x)F(x)s}{(1-F(x)s)^2} dx \right) \times \left(-\frac{1}{s} \log(1-s) \right)^{m-1} \right\rangle^{(n-m)}.
 \end{aligned}$$

عبارت $\frac{1}{(1-F(x)s)^2}$ را با استفاده از بسط سری های توانی، می توان به صورت زیر نوشت:

$$\frac{1}{(1-F(x)s)^2} = \sum_{i=0}^{+\infty} (i+1) F^i(x) s^i.$$

بنابراین برای $A_2(m)$ داریم

$$A_2(m) = \left\langle \sum_{i=0}^{+\infty} \left(s^i \int_{-\infty}^{+\infty} (-\ln F(x)) i F^i(x) f(x) dx \right) \times \frac{1}{(m-1)!} \left(-\frac{1}{s} \log(1-s) \right)^{m-1} \right\rangle^{(n-m)}.$$

اکنون با قرار دادن $A_2(m)$ در A_2 و با تغییر ترتیب مجموع ها، و با استدلال های مشابه قبل،

به دست می آوریم

$$A_2 = \sum_{i=1}^{n-1} i \int_{-\infty}^{+\infty} (-\ln F(x)) F^i(x) f(x) dx. \quad (۸-۲)$$

فصل دوم: اطلاع در داده های رکوردی

حال با جایگذاری (۲-۸) و (۲-۶) در (۲-۵)، (۲-۴) حاصل می شود و برهان کامل می شود. □

تعریف ۲-۱-۲ اگر x یک مقدار باشد تابع $\psi(x)$ به صورت زیر تعریف می شود

$$\psi(x) = \frac{d}{dx} \ln(\Gamma(x)) = \frac{\frac{d}{dx}(\Gamma(x))}{\Gamma(x)}.$$

نتیجه ۲-۱-۳ (***) با استفاده از قضیه ۲-۱-۱، داریم

$$H_{RT}^U(n) = \psi(n+1) + \gamma - \sum_{i=1}^n \frac{1}{i^2} - \sum_{i=0}^{n-1} \phi_f^U(i) \quad (9-2)$$

که $\phi_f^U(i)$ را با استفاده از (۲-۲) می توان به دست آورد، $\gamma = 0.57721566$ ثابت اوایلر و ψ

نیز تابع دی گاما^{۶۶} می باشد.

برهان. از رابطه (۲-۴) داریم

$$\begin{aligned} H_{RT}^U(n) &= \sum_{i=0}^{n-1} \int_{-\infty}^{+\infty} (-\ln f(x) - i \ln F(x)) F^i(x) f(x) dx \\ &= \sum_{i=0}^{n-1} \left[\int_{-\infty}^{+\infty} (-\ln f(x)) F^i(x) f(x) dx + \frac{i}{(i+1)^2} \right] \\ &= \sum_{i=1}^n \frac{i-1}{i^2} - \sum_{i=0}^{n-1} \int_{-\infty}^{+\infty} (\ln f(x)) F^i(x) f(x) dx. \end{aligned}$$

اما

$$\sum_{i=0}^n \frac{i-1}{i^2} = \sum_{i=1}^n \frac{1}{i} - \sum_{i=1}^n \frac{1}{i^2} = \psi(n+1) + \gamma - \sum_{i=1}^n \frac{1}{i^2}.$$

بنابراین رابطه (۲-۹) حاصل می شود. □

⁶⁶ digamma

فصل دوم: اطلاع در داده های رکوردی

برای مقادیر رکورد و زمان های رکورد پایین نیز به طور مشابه داریم

$$H_{RT}^L(n) = \psi(n+1) + \gamma - \sum_{i=1}^n \frac{1}{i^2} - \sum_{i=0}^{n-1} \phi_f^L(i). \quad (10-2)$$

که در آن $\phi_f^L(i)$ از $\phi_f^U(i)$ با جایگذاری F با $1-F$ بدست می آید.

نتیجه ۲-۱-۴ با استفاده از (۱-۲) و (۴-۲) می توانیم نشان دهیم

$$H_{RT}^U(n) = \sum_{i=1}^n \frac{1}{i} (\ln i + H_M^U(i)). \quad (11-2)$$

برای مشاهده برهان به [۲] مراجعه شود.

نتیجه ۲-۱-۵ فرض کنید $Z_1, Z_2, \dots, Z_n$ مستقل و هم توزیع با تابع توزیع تجمعی F_Z و تابع

چگالی احتمال f_Z باشند و اطلاع شانون متناظر با مقادیر و زمان های رکورد بالایی با

$H_{RT}^U(n, Z)$ مشخص شود. می توان نشان داد [۲].

۱. اگر $X = Z + \theta$ ، آن گاه به ازای هر $\theta \in \mathbb{R}$ داریم

$$H_{RT}^U(n, X) = H_{RT}^U(n, Z).$$

۲. اگر $X = \theta Z$ ، آن گاه به ازای هر $\theta > 0$ داریم

$$H_{RT}^U(n, X) = H_{RT}^U(n, Z) + (\psi(n+1) + \gamma) \ln \theta.$$

۳. اگر $X = Z^\theta$ ، آن گاه به ازای هر $\theta > 0$ داریم

$$H_{RT}^U(n, X) = H_{RT}^U(n, Z) + (\psi(n+1) + \gamma) \ln \theta \\ + (\theta - 1) \sum_{i=0}^{n-1} \int_{-\infty}^{+\infty} (\ln z) F_Z^i(z) f_Z(z) dz.$$

فصل دوم: اطلاع در داده های رکوردی

به طور مشابه، تمام روابط بالا برای آنتروپی مقادیر و زمان های رکورد پایینی در نمونه ای با توابع توزیع F_x و F_z ، برقرار هستند.

۲-۱-۲- اطلاع در مقادیر رکورد از یک طرح نمونه گیری معکوس:

در طرح نمونه گیری معکوس، عمل نمونه گیری تا زمانی انجام می شود که به یک مقدار ثابت m از رکوردها دست یابیم. برای مقدار ثابت m ، احتمال توأم رکوردهای بالا، $R_1, \dots, R_m$ ، و زمانهای رکورد، $T_1, \dots, T_m$ ، از (۴-۱) به دست می آید. تابع چگالی توأم مقادیر رکورد $R_1, \dots, R_m$ نیز از (۵-۱) و تابع چگالی احتمال حاشیه ای آخرین رکورد از (۶-۱) بدست می آید.

با توجه به این مطالب، می توان اطلاع شانون را برای این نوع رکورد به دست آورد. فرض کنید $h_M^U(m)$ ، $h_R^U(m)$ و $h_{RT}^U(m)$ به ترتیب اطلاع شانون آخرین رکورد بالایی، مقادیر رکورد بالایی و اطلاع شانون مقادیر و زمان های رکورد بالایی باشند. در این صورت می توان قضیه زیر را مطرح کرد.

تعریف ۶-۱-۲ گر x یک مقدار باشد، تابع $\zeta(x)$ به صورت زیر تعریف می شود

$$\zeta(x) = \sum_{i=1}^{\infty} \frac{1}{i^x}.$$

قضیه ۷-۱-۲ (مددی و تاتا ۲۰۱۱) اطلاع شانون آخرین رکورد بالایی، مقادیر رکورد بالایی و مقادیر و زمان های رکورد بالایی در طرح نمونه گیری معکوس، به صورت زیر به دست می آیند [۳۰]

$$h_M^U(m) = \sum_{i=1}^{m-1} \ln i - (m-1)\psi(m) - \psi_f^U(m), \quad (12-2)$$

$$h_R^U(m) = -\sum_{i=1}^m \psi_f^U(i) + \frac{m(m-1)}{2}, \quad (13-2)$$

$$h_{RT}^U(m) = h_R^U(m) - \sum_{i=2}^m (m-i+1) \zeta(i), \quad (۱۴-۲)$$

که در تمام روابط بالا، $\psi_f^U(i)$ به صورت

$$\begin{aligned} \psi_f^U(i) &= E(\ln f(R_i)) \\ &= \frac{1}{\Gamma(i)} \int_{-\infty}^{+\infty} (\ln f(x)) \times (-\ln(1-F(x)))^{i-1} f(x) dx, \end{aligned} \quad (۱۵-۲)$$

تعریف می شود و همچنین $h_M^U(1) = h_{RT}^U(1) = -\psi_f^U(1)$ ، تابع دی گاما و ζ تابع زتا^{۶۷} می باشد.

برهان: عبارت (۱۲-۲) به راحتی از تعریف اطلاع شانون و تابع چگالی (۶-۱) به دست می آید. برای مشاهده جزئیات اثبات روابط بالا، می توان به [۲۸] مراجعه کرد. با استفاده از رابطه (۵-۱)، اطلاع شانون در رکوردهای بالا را می توان به صورت زیر نوشت:

$$\begin{aligned} h_R^U(m) &= E(-\ln f(R_1, \dots, R_m)) = E\left(-\sum_{i=1}^m \ln f(R_i)\right) - \sum_{i=1}^{m-1} E(\ln(1-F(R_i))) \\ &= -\sum_{i=1}^m \psi_f^U(i) - \sum_{i=1}^{m-1} \frac{1}{\Gamma(i)} \int_{-\infty}^{+\infty} (\ln(1-F(x))) \times (-\ln(1-F(x)))^{i-1} f(x) dx. \end{aligned} \quad (۱۶-۲)$$

حالا با قرار دادن $u = -\ln(1-F(x))$ ، رابطه (۱۳-۲) حاصل می شود. با استفاده از (۴-۱)، اطلاع شانون مقادیر و زمان های رکورد بالایی را می توان به صورت زیر نوشت:

$$\begin{aligned} h_{RT}^U(m) &= E(-\ln f(R_1, \dots, R_m, T_1, \dots, T_m)) \\ &= \sum_{i=1}^m E(-\ln f(R_i)) - \sum_{i=1}^{m-1} E(\delta_i \ln F(R_i)) \end{aligned} \quad (۱۷-۲)$$

⁶⁷Zeta

$$\begin{aligned}
 &= \sum_{i=1}^m E(-\ln f(R_i)) - \sum_{i=1}^{m-1} E[E(\delta_i \ln F(R_i) | R_i)] \\
 &= \sum_{i=1}^m E(-\ln f(R_i)) - \sum_{i=1}^{m-1} E[\ln F(R_i) \cdot E(\delta_i | R_i)].
 \end{aligned}$$

همچنین در زمان $\delta_i + 1$ ، هنگامی که شرط روی R_i می باشد، ورود i امین رکورد دارای توزیع هندسی با پارامتر $(1 - F(R_i))$ است. پس مقدار $E(\delta_i | R_i)$ به صورت زیر به دست می آید:

$$E(\delta_i | R_i) = \frac{F(R_i)}{1 - F(R_i)}.$$

حال با جایگذاری در رابطه (۲-۱۷) مقدار $h_{RT}^U(m)$ به صورت زیر حاصل می شود:

$$\begin{aligned}
 h_{RT}^U(m) &= -\sum_{i=1}^m \psi_f^U(i) + \sum_{i=1}^{m-1} E\left(\frac{F(R_i)}{1 - F(R_i)} \ln(F(R_i))\right) \\
 &= -\sum_{i=1}^m \psi_f^U(i) - \sum_{i=1}^{m-1} \frac{1}{\Gamma(i)} \int_{-\infty}^{+\infty} \frac{F(x)}{1 - F(x)} (-\ln F(x)) \times (-\ln(1 - F(x)))^{i-1} f(x) dx.
 \end{aligned}$$

با تغییر متغیر $u = -\ln(1 - F(x))$ داریم

$$h_{RT}^U(m) = -\sum_{i=1}^m \psi_f^U(i) - \sum_{i=1}^{m-1} \frac{1}{\Gamma(i)} \int_0^{+\infty} (1 - e^{-u}) \times (-\ln(1 - e^{-u})) u^{i-1} du.$$

از آنجایی که برای $|s| < 1$ ، $-\ln(1 - s) = \sum_{k=1}^{+\infty} \frac{s^k}{k}$ ، به صورت زیر حاصل می شود:

$$h_{RT}^U(m) = -\sum_{i=1}^m \psi_f^U(i) - \sum_{i=1}^{m-1} \frac{1}{\Gamma(i)} \int_0^{+\infty} (1 - e^{-u}) \times \left(\sum_{k=1}^{+\infty} \frac{e^{-ku}}{k} \right) u^{i-1} du.$$

اما

$$(1 - e^{-u}) \sum_{k=1}^{+\infty} \frac{e^{-ku}}{k} = e^{-u} - \sum_{k=2}^{+\infty} \left(\frac{1}{k-1} - \frac{1}{k} \right) e^{-ku}.$$

بنابراین

$$\begin{aligned}
 h_{RT}^U(m) &= -\sum_{i=1}^m \psi_f^U(i) - \sum_{i=1}^{m-1} \frac{1}{\Gamma(i)} \int_0^{+\infty} \left(e^{-u} - \sum_{k=2}^{+\infty} \frac{e^{-ku}}{k(k-1)} \right) u^{i-1} du \\
 &= -\sum_{i=1}^m \psi_f^U(i) - \sum_{i=1}^{m-1} \frac{1}{\Gamma(i)} \left[\Gamma(i) - \sum_{k=2}^{+\infty} \frac{1}{k(k-1)} \cdot \frac{\Gamma(i)}{k^i} \right] \\
 &= -\sum_{i=1}^m \psi_f^U(i) - (m-1) + \sum_{i=1}^{m-1} \sum_{k=2}^{+\infty} \frac{1}{(k-1)k^{i+1}}. \quad (18-2)
 \end{aligned}$$

اکنون

$$\begin{aligned}
 \sum_{k=2}^{+\infty} \frac{1}{(k-1)k^{i+1}} &= \sum_{k=2}^{+\infty} \left[\frac{1}{(k-1)} - \frac{k^i + k^{i-1} + \dots + k + 1}{k^{i+1}} \right] \\
 &= \sum_{k=2}^{+\infty} \left[\left(\frac{1}{(k-1)} - \frac{1}{k} \right) - \left(\frac{1}{k^2} + \frac{1}{k^3} + \dots + \frac{1}{k^{i+1}} \right) \right] \\
 &= 1 + i - \sum_{k=1}^{+\infty} \left(\frac{1}{k^2} + \frac{1}{k^3} + \dots + \frac{1}{k^{i+1}} \right) \\
 &= 1 + i - \sum_{k=1}^{+\infty} \left(\sum_{j=2}^{i+1} \frac{1}{k^j} \right)
 \end{aligned}$$

با تغییر نظم در مجموع داریم

$$\begin{aligned}
 &= 1 + i - \sum_{j=2}^{i+1} \sum_{k=1}^{+\infty} \left(\frac{1}{k^j} \right) \\
 &= 1 + i - \sum_{j=2}^{i+1} \zeta(j). \quad (19-2)
 \end{aligned}$$

با جانشین کردن (19-2) در (18-2)، مقدار $h_{RT}^U(m)$ حاصل و برهان کامل می شود. □

باید توجه داشته باشید که برای مقادیر رکورد پایینی نیز به طور مشابه، داریم

$$h_M^L(m) = \sum_{i=1}^{m-1} \ln i - (m-1)\psi(m) - \psi_f^L(m), \quad (20-2)$$

$$h_R^L(m) = -\sum_{i=1}^m \psi_f^L(i) + \frac{m(m-1)}{2}, \quad (21-2)$$

$$h_{RT}^L(m) = h_R^L(m) - \sum_{i=2}^m (m-i+1)\zeta(i), \quad (22-2)$$

به طوری که، $\psi_f^L(i)$ از $\psi_f^U(i)$ با جایگذاری F با $1-F$ حاصل می شود.

فرض کنید $\{Z_i, i=1,2,\dots\}$ دنباله های از متغیر های هم توزیع و مستقل با تابع توزیع احتمال

F_Z و تابع چگالی احتمال f_Z باشند و مقدار اطلاع شانون در آخرین رکورد و مقادیر رکورد بالا

به ترتیب متناظر با $h_M^U(m, Z)$ و $h_R^U(m, Z)$ باشد. می توان نشان داد

• اگر $X = Z + \theta$ ، آن گاه به ازای هر $\theta \in \mathbb{R}$ داریم

$$h_M^U(m, X) = h_M^U(m, Z)$$

$$h_R^U(m, X) = h_R^U(m, Z)$$

• اگر $X = \theta Z$ ، آن گاه به ازای هر $\theta > 0$ داریم

$$h_M^U(m, X) = h_M^U(m, Z) + \ln \theta$$

$$h_R^U(m, X) = h_R^U(m, Z) + m \ln \theta$$

• اگر $X = Z^\theta$ ، آن گاه به ازای هر $\theta > 0$ داریم

$$h_M^U(m, X) = h_M^U(m, Z) + \ln \theta$$

$$+ \frac{\theta-1}{\Gamma(m)} \int_{-\infty}^{+\infty} (\ln z) \times (-\ln(1-F_Z(z)))^{m-1} f_Z(z) dz,$$

$$h_R^U(m, X) = h_R^U(m, Z) + m \ln \theta + (\theta-1)$$

$$\times \sum_{i=1}^{m-1} \frac{1}{\Gamma(i)} \int_{-\infty}^{+\infty} (\ln z) \times (-\ln(1-F_Z(z)))^{m-1} f_Z(z) dz,$$

فصل دوم: اطلاع در داده های رکوردی

تمام روابط بالا نیز، به طور مشابه برای اطلاع شانون مقادیر و زمان های رکورد پایینی در نمونه ای با تابع توزیع F_x و F_z برقرار هستند. در ادامه، مقدار اطلاع شانون را در مدل کلاسیک و طرح نمونه گیری معکوس برای دو توزیع معروف یکنواخت و نمایی به دست می آوریم [۲].

مثال ۱-۱-۲ توزیع یکنواخت

فرض کنید X دارای توزیع یکنواخت $(0,1)$ ، با تابع چگالی $f_x(x) = I_{(0,1)}(x)$ باشد که در آن تابع $I_A(\cdot)$ نشانگر استو به صورت زیر تعریف می شود،

$$I_A(x) = \begin{cases} 1 & x \in A, \\ 0 & x \notin A. \end{cases}$$

با توجه به این که توزیع حول $\frac{1}{2}$ متقارن می باشد، بنابراین با استفاده از (۲-۲) و (۱۵-۲) داریم

$$\phi_f^U(i) = \int_0^1 (\ln f(x)) F^i(x) f(x) dx = \int_0^1 (\ln 1) F^i(x) dx = 0 \quad i = 1, 2, 3, \dots,$$

$$\begin{aligned} \psi_f^U(i) &= \frac{1}{\Gamma(i)} \int_0^1 (\ln f(x)) \times (-\ln(1-F(x)))^{i-1} f(x) dx \\ &= \frac{1}{\Gamma(i)} \int_0^1 (\ln 1) \times (-\ln(1-F(x)))^{i-1} dx = 0 \quad i = 1, 2, 3, \dots \end{aligned}$$

با توجه به خاصیت تقارن و با استفاده از (۱-۲) و (۳-۲)، داریم

$$H_M^U(n) = H_M^L(n) = -\ln n + \frac{n-1}{n}.$$

با استفاده از (۹-۲) و (۱۰-۲)، می توان نوشت

$$H_{RT}^U(n) = H_{RT}^L(n) = \psi(n+1) + \gamma - \sum_{i=1}^n \frac{1}{i^2}.$$

فصل دوم: اطلاع در داده های رکوردی

برای اطلاع شانون در طرح نمونه‌گیری معکوس از توزیع یکنواخت، با استفاده از (۲-۱۲)،

(۲-۱۳)، (۲-۱۴) و (۲-۲۰)، (۲-۲۱) و (۲-۲۲)، داریم

$$h_M^U(m) = h_M^L(m) = \sum_{i=1}^{m-1} \ln i - (m-1)\psi(m),$$

$$h_R^U(m) = h_R^L(m) = \frac{m(m-1)}{2},$$

$$h_{RT}^U(m) = h_{RT}^L(m) = \frac{m(m-1)}{2} - \sum_{i=2}^m (m-i+1)\zeta(i).$$

جدولهای ۲-۱ و ۲-۲، مقدار اطلاع شانون در مدل کلاسیک و طرح نمونه‌گیری معکوس را برای

توزیع یکنواخت، به ازای مقادیر $n = 5, 10, 20$ و $m = 5, 10, 20$ ، نشان می‌دهند.

جدول ۲-۱: اطلاع شانون در مدل کلاسیک برای توزیع یکنواخت

n	$H_M^U(n) = H_M^L(n)$	$H_{RT}^U(n) = H_{RT}^L(n)$
۵	-۰٫۸	۰٫۸۲
۱۰	-۱٫۴	۱٫۳۸
۲۰	-۲٫۰۵	۲٫۰

جدول ۲-۲: اطلاع شانون در طرح نمونه‌گیری معکوس برای توزیع یکنواخت

m	$h_M^U(m) = h_M^L(m)$	$h_R^U(m) = h_R^L(m)$	$h_{RT}^U(m) = h_{RT}^L(m)$
۵	-۲٫۸۵	۱۰	-۳٫۳۹
۱۰	-۷٫۴۶	۴۵	-۸٫۳۶
۲۰	-۱۷٫۱۰	۱۹۰	-۱۸٫۳۹

فصل دوم: اطلاع در داده های رکوردی

همان طور که از جداول مشاهده می شود با افزایش تعداد رکوردها مقدار آنتروپی شانون افزایش می یابد، که مورد انتظار است.

مثال ۲-۱-۲ توزیع نمایی

فرض کنید X دارای توزیع نمایی با پارامتر ۱ و تابع چگالی احتمال $f_X(x) = e^{-x} I_{(0,+\infty)}(x)$

باشد. با استفاده از (۲-۲) داریم

$$\phi_f^U(i) = \int_0^{+\infty} (-x)(1-e^{-x})^i e^{-x} dx.$$

با تغییر متغیر $1-e^{-x} = u$ و جایگذاری $-\ln(1-u) = \sum_{k=1}^{+\infty} \frac{u^k}{k}$ داریم

$$\begin{aligned} \phi_f^U(i) &= -\int_0^{+\infty} \ln(1-u) u^i dx \\ &= -\int_0^1 \left(\sum_{k=1}^{+\infty} \frac{u^k}{k} \right) u^i du \\ &= -\sum_{k=1}^{+\infty} \frac{1}{k} \int_0^1 u^{k+i} du \\ &= -\sum_{k=1}^{+\infty} \frac{1}{k(k+i+1)} = -\frac{1}{i+1} \sum_{k=1}^{i+1} \frac{1}{k} \\ &= -\frac{1}{i+1} (\psi(i+2) + \gamma). \end{aligned}$$

9

$$\begin{aligned} \phi_f^U(i) &= \int_0^1 (\ln f(x))(1-F(x))^i f(x) dx \\ &= \int_0^1 (-x)(e^{-x})^i e^{-x} dx \\ &= -\frac{\Gamma(2)}{(i+1)^2}. \end{aligned}$$

بنابراین

$$H_M^U(n) = -\ln n + \frac{n-1}{n} + (\psi(n+1) + \gamma),$$

$$H_M^L(n) = -\ln n + \frac{n-1}{n} + \frac{1}{n}.$$

با استفاده از (۹-۲) و (۱۰-۲)، $H_{RT}^U(n)$ و $H_{RT}^L(n)$ به صورت زیر حاصل می شوند

$$H_{RT}^U(n) = (\gamma + 1)(\psi(n+1) + \gamma) - \sum_{i=1}^n \frac{\psi(n+1)}{i} - \sum_{i=1}^n \frac{1}{i^2},$$

$$H_{RT}^L(n) = \psi(n+1) + \gamma - \sum_{i=1}^n \frac{1}{i^2} + \sum_{i=1}^n \frac{1}{i^3}.$$

برای محاسبه اطلاع شانون توزیع نمایی در طرح نمونه گیری معکوس، با استفاده از (۱۵-۲) داریم

$$\begin{aligned} \psi_f^U(i) &= \frac{1}{\Gamma(i)} \int_0^{+\infty} (-x) \times (x)^{i-1} e^{-x} dx \\ &= -\frac{1}{\Gamma(i)} \int_0^{+\infty} x^i e^{-x} dx \\ &= -\frac{\Gamma(i+1)}{\Gamma(i)} \\ &= -i. \end{aligned}$$

و

$$\psi_f^L(i) = \frac{1}{\Gamma(i)} \int_0^{+\infty} (-x) \times (-\ln(1-e^{-x}))^{i-1} e^{-x} dx.$$

با تغییر متغیر $1-e^{-x} = u$ داریم

$$\psi_f^L(i) = -\frac{1}{\Gamma(i)} \int_0^1 \left(\sum_{k=1}^{+\infty} \frac{u^k}{k} \right) (-\ln u)^{i-1} du.$$

با تغییر متغیر $-\ln u = x$ داریم

$$\begin{aligned}\psi_f^L(i) &= \sum_{k=2}^{+\infty} \frac{1}{k^i (k-1)} \\ &= -\left(i - \sum_{j=2}^i \zeta(j)\right) \quad \forall i > 1.\end{aligned}$$

که در آن، $\psi_f^L(1) = -1$ حال با جایگزینی $\psi_f^U(i)$ در $(12-2)$ ، $(13-2)$ و $(14-2)$ داریم

$$h_M^U(m) = \sum_{I=1}^{m-1} \ln i - (m-1)\psi(m) + m,$$

$$h_R^U(m) = m,$$

$$h_{RT}^U(m) = m - \sum_{i=2}^m (m-i+1)\zeta(i).$$

و برای مقادیر رکورد پایینی، با جایگزینی $\psi_f^L(i)$ در $(20-2)$ ، $(21-2)$ و $(22-2)$ داریم

$$h_M^L(m) = \sum_{I=1}^{m-1} \ln i - (m-1)\psi(m) + m - \sum_{I=2}^m \zeta(i),$$

$$h_{RT}^L(m) = m - \sum_{i=2}^m (m-i+1)\zeta(i),$$

$$h_{RT}^L(m) = m.$$

جدولهای ۲-۳ و ۲-۴، مقدار اطلاع شانون در مدل کلاسیک و طرح نمونه گیری معکوس را برای

توزیع نمایی، به ازای مقادیر $n = 5, 10, 20$ و $m = 5, 10, 20$ ، نمایش می دهند.

فصل دوم: اطلاع در داده های رکوردی

جدول ۲-۳: اطلاع شانون در مدل کلاسیک برای توزیع نمایی

n	$H_M^L(n)$	$H_{RT}^L(n)$	$H_M^U(n)$	$H_{RT}^U(n)$
۵	-۰/۶۱	۲/۲۸	۱/۴۷	۴/۱۶
۱۰	-۱/۳۰	۲/۹۳	۱/۵۳	۶/۴۴
۲۰	-۲/۰	۳/۶۰	۱/۵۵	۹/۲۷

جدول ۲-۴: اطلاع شانون در طرح نمونه گیری معکوس برای توزیع نمایی

m	$h_M^L(m)$	$h_R^L(m)$	$h_{RT}^L(m)$	$h_M^U(m)$	$h_R^U(m)$	$h_{RT}^U(m)$
۵	-۲/۷۱	-۸/۳۹	۵	۲/۱۵	۵	۱۸/۳۹
۱۰	-۷/۴۶	-۴۳/۳۶	۱۰	۲/۵۴	۱۰	۶۳/۳۶
۲۰	-۱۷/۱۰	-۱۸۸/۳۶	۲۰	۲/۹۰	۲۰	۲۲۸/۳۶

همانطور که از مقادیر جداول بالا مشاهده می شود، با افزایش تعداد رکوردها مقدار آنترپی شانون افزایش می یابد.

فصل ۳

اطلاع شانون مقادیر و زمان

رکوردها در توزیع t – استودنت

۳-۱- مقدمه

در این فصل، ابتدا توزیع t -استودنت را معرفی نموده، سپس آنتروپی توزیع t -استودنت و F را محاسبه می‌کنیم و در انتها، اطلاع شانون مقادیر و زمان رکوردها در توزیع t -استودنت را مورد بررسی قرار می‌دهیم.

۳-۱-۱- توزیع t -استودنت

ویلیام سیلی گوسه^{۶۸} در ۱۳ ژوئن^{۶۹} ۱۸۷۶ در کانتربری^{۷۰} انگلستان به دنیا آمد و در کالج وینچستر^{۷۱} به تحصیل پرداخت. سپس در نیو کالج^{۷۲} آکسفورد^{۷۳} به تحصیل شیمی و ریاضیات مشغول شد، اما رشته اصلی فعالیت او آمار بود. او یکی از آماردانان مشهور زمان خویش بود که نمی‌توانست به نام اصلیش مقالاتش را امضاء کند و به ناچار از نام مستعار استودنت (شاگرد) استفاده کرد و کارهایش را در زمینه توزیع t -استودنت منتشر کرد.

William Sealy Gosset



Student in 1908

در هنگام تعیین تقریبی میانگین نمونه‌های برداشته شده از یک متغیر تصادفی، توزیع t -استودنت مطرح می‌شود. این توزیع اساس آزمونی به نام «آزمون t » است که میزان اطمینان از تفاوت دو متغیر تصادفی را از روی نمونه‌هایشان اعلام می‌کند. آزمون t -استودنت برای ارزیابی میزان هم‌قوارگی یا یکسان بودن و نبودن میانگین نمونه‌ای با میانگین جامعه، در حالتی به کار می‌رود که انحراف معیار

⁶⁸ William Sealy Gosset

⁶⁹ June

⁷⁰ Canterbury

⁷¹ Winchester

⁷² New College

⁷³ Oxford

جامعه نامعلوم باشد، چون توزیع t -استودنت در نمونه‌های کوچک با کمک درجات آزادی تعدیل می‌شود، می‌توان از این آزمون برای نمونه‌های بسیار کوچک استفاده کرد.

می‌گوییم متغیر تصادفی T دارای توزیع t با r درجه آزادی است و با نماد $T \sim t_{(r)}$ نشان می‌دهیم اگر دارای تابع چگالی به صورت زیر باشد

حالت اول:

$$f_T(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{t^2}{r}\right)^{-\left(\frac{r+1}{2}\right)} \quad -\infty < t < +\infty \quad (1-3)$$

که در آن $\Gamma(\cdot)$ تابع گاما می‌باشد و عبارتست از

$$\Gamma(\alpha) = \int_0^{+\infty} y^{\alpha-1} e^{-y} dy \quad (2-3)$$

که در آن، $\alpha > 0$.

حالت دوم:

$$f_T(t) = \frac{1}{\sqrt{r}B\left(\frac{1}{2}, \frac{r}{2}\right)} \left(1 + \frac{t^2}{r}\right)^{-\left(\frac{r+1}{2}\right)} \quad -\infty < t < +\infty \quad (3-3)$$

که در آن $B(\cdot, \cdot)$ تابع بتا می‌باشد و به صورت زیر تعریف می‌شود:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx \quad \alpha > 0, \beta > 0 \quad (4-3)$$

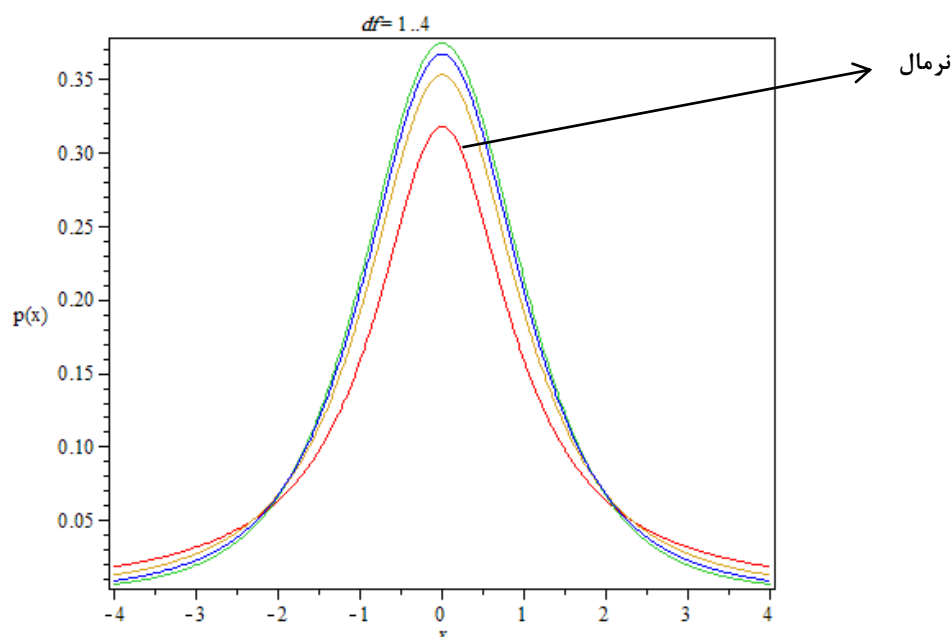
و معادل است با $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

حال برخی از خواص توزیع t -استودنت را بررسی می‌کنیم. اگر X دارای توزیع t با r درجه آزادی باشد، آن‌گاه

- میانگین توزیع t -استودنت، برای $r > 1$ برابر صفر است و برای $r = 1$ تعریف نشده است.
- واریانس توزیع t -استودنت، برای $r > 2$ برابر $\frac{r}{r-2}$ می‌باشد.
- تابع مولد گشتاور^{۷۴} توزیع t -استودنت تعریف نشده است.
- گشتاور مرکزی k ام توزیع t -استودنت برابر است با

$$E(t^k) = \begin{cases} 0 & 0 < k < r, \quad k \text{ زوج} \\ \frac{1}{\sqrt{\pi} \Gamma\left(\frac{r}{2}\right)} \left[\Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{r-k}{2}\right) r^{\frac{k}{2}} \right] & 0 < k < r, k \text{ فرد} \end{cases}$$

- توزیع t -استودنت از خانواده توزیع‌های پیوسته و حول صفر متقارن می‌باشد، به این معنی که برای آن، $P(-x \leq t < 0) = P(0 < t \leq x)$ در شکل ۱-۳ تابع چگالی t -استودنت برای چند درجه آزادی متفاوت نشان داده شده‌اند که این واقعیت را نشان می‌دهد.



شکل ۱-۳: نمودار تابع چگالی احتمال t -استودنت و نرمال

⁷⁴ Moment Generating Function

همان طور که از شکل ۱-۳ دیده می شود، با افزایش درجه آزادی توزیع t به توزیع نرمال نزدیک می شود.

- می توان توزیع t -استودنت را به کمک توزیع نرمال تقریب زد به طوری که

$$p(t_r \leq t) \approx p\left(Z \leq \frac{t\left(1 - \frac{1}{4r}\right)}{\sqrt{1 + \frac{t^2}{2r}}}\right),$$

که در آن، Z متغیر تصادفی نرمال استاندارد می باشد.

۳-۱-۲- روش تولید داده تصادفی از توزیع t -استودنت

تولید داده تصادفی از توزیع t -استودنت در سه مرحله انجام می گیرد.

(۱) متغیر تصادفی Z را از $N(0,1)$ تولید می کنیم.

(۲) متغیر تصادفی S را از $\Gamma\left(\frac{r}{2}, 2\right)$ تولید می کنیم.

(۳) قرار می دهیم

$$x = \frac{Z}{\sqrt{\frac{S}{r}}}$$

که x یک متغیر تصادفی t -استودنت با درجه آزادی r می باشد.

۳-۱-۳- تابع توزیع تجمعی t -استودنت

تابع توزیع تجمعی t -استودنت را می توان با استفاده از تعریف تابع توزیع تجمعی، به صورت زیر نوشت [۳۱]:

$$F_T(t) = \int_{-\infty}^t f(u) du = \frac{1}{2} + t \times \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) \times {}_2F_1\left(\frac{1}{2}, \frac{r+1}{2}; \frac{3}{2}; -\frac{t^2}{r}\right), \quad (5-3)$$

که به ازای هر $t^2 < r$ برقرار است و در آن، تابع فوق هندسی است و به صورت زیر تعریف می شود:

تعریف ۳-۱-۱ تابع فوق هندسی برای $|z| < 1$ با استفاده از سری های توانی به صورت زیر تعریف می شود:

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \times \frac{z^n}{n!}, \quad a, b, c \neq -1, -2, \dots, \quad c \neq 0,$$

که در آن، $(q)_n$ برابر است با

$$(q)_n = \begin{cases} 1 & n = 0, \\ q(q+1)\dots(q+n+1) & n > 0. \end{cases}$$

۳-۱-۴- آنروپی توزیع t -استودنت

در این بخش، آنروپی توزیع t -استودنت را در قالب یک قضیه مطرح می کنیم.

قضیه ۳-۱-۲ ()** فرض کنید متغیر تصادفی X دارای توزیع t با r درجه آزادی باشد، در

این صورت آنروپی این توزیع عبارت است از

$$H(X) = \ln\left(\sqrt{r} \cdot B\left(\frac{1}{2}, \frac{r}{2}\right)\right) + \frac{r+1}{2} \cdot \left[\psi\left(\frac{r+1}{2}\right) - \psi\left(\frac{r}{2}\right)\right]. \quad (6-3)$$

برهان. با استفاده از (۱۲-۱) و (۳-۳) داریم

$$H(X) = -\int_{-\infty}^{+\infty} f(x) \ln f(x) dx = -\int_{-\infty}^{+\infty} f(x) \ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\left(\frac{r+1}{2}\right)} \right) dx$$

$$\begin{aligned}
 &= -\int_{-\infty}^{+\infty} f(x) \ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) dx - \int_{-\infty}^{+\infty} f(x) \ln \left(1 + \frac{x^2}{r} \right)^{-\left(\frac{r+1}{2}\right)} dx \\
 &= -\ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) \cdot \int_{-\infty}^{+\infty} f(x) dx + \left(\frac{r+1}{2} \right) \int_{-\infty}^{+\infty} f(x) \ln \left(1 + \frac{x^2}{r} \right) dx.
 \end{aligned}$$

می‌دانیم $\int_{-\infty}^{+\infty} f(x) dx$ برابر است با ۱، بنابراین

$$H(X) = -\ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) + \left(\frac{r+1}{2} \right) \cdot \frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \int_{-\infty}^{+\infty} \left(1 + \frac{x^2}{r} \right)^{-\left(\frac{r+1}{2}\right)} \cdot \ln \left(1 + \frac{x^2}{r} \right) dx$$

از آنجایی که توزیع t-استودنت یک توزیع متقارن^{۷۵} است، با تغییر متغیر $1 + \frac{x^2}{r} = u^2$ و

جایگذاری، داریم

$$\begin{aligned}
 H(X) &= -\ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) + \left(\frac{r+1}{2} \right) \cdot \frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \cdot 2\sqrt{r} \int_1^{+\infty} \frac{\ln(u)}{u^r \sqrt{u^2 - 1}} du \\
 &= -\ln \left(\frac{1}{\sqrt{r} \cdot B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) + \frac{r+1}{B\left(\frac{1}{2}, \frac{r}{2}\right)} \int_1^{+\infty} \frac{\ln(u)}{u^r \sqrt{u^2 - 1}} du \\
 &= \ln \left(\sqrt{r} \cdot B\left(\frac{1}{2}, \frac{r}{2}\right) \right) + \frac{r+1}{B\left(\frac{1}{2}, \frac{r}{2}\right)} \int_1^{+\infty} \frac{\ln(u)}{u^r \sqrt{u^2 - 1}} du.
 \end{aligned}$$

⁷⁵ Symmetric

بنابراین

$$\begin{aligned} H(X) &= \ln \left(\sqrt{r} \cdot B \left(\frac{1}{2}, \frac{r}{2} \right) \right) + \frac{r+1}{B \left(\frac{1}{2}, \frac{r}{2} \right)} \cdot \left(\frac{1}{2} \cdot \frac{\sqrt{\pi} \Gamma \left(\frac{r}{2} \right) \left[\psi \left(\frac{r+1}{2} \right) - \psi \left(\frac{r}{2} \right) \right]}{\Gamma \left(\frac{r+1}{2} \right)} \right) \\ &= \ln \left(\sqrt{r} \cdot B \left(\frac{1}{2}, \frac{r}{2} \right) \right) + \frac{r+1}{B \left(\frac{1}{2}, \frac{r}{2} \right)} \cdot \frac{B \left(\frac{1}{2}, \frac{r}{2} \right)}{2} \cdot \left[\psi \left(\frac{r+1}{2} \right) - \psi \left(\frac{r}{2} \right) \right] \\ &= \ln \left(\sqrt{r} \cdot B \left(\frac{1}{2}, \frac{r}{2} \right) \right) + \frac{r+1}{2} \left[\psi \left(\frac{r+1}{2} \right) - \psi \left(\frac{r}{2} \right) \right]. \end{aligned}$$

و برهان کامل می‌شود. □

قضیه زیر نشان می‌دهد که با داشتن آنتروپی یک متغیر تصادفی، می‌توان آنتروپی توان دوم آن را نیز بدست آورد.

قضیه ۳-۱-۳ (*)** اگر X یک متغیر تصادفی با تابع چگالی متقارن $f_X(x)$ و آنتروپی $H(X)$ باشد، در این صورت آنتروپی $Y = X^2$ عبارت‌است از

$$H(Y) = 2 \int_0^{+\infty} \ln(u) \cdot f_X(x) du + H(X). \quad (۷-۳)$$

برهان: می‌دانیم Y دارای تابع چگالی

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}), \quad 0 < y < +\infty,$$

است، بنابراین با استفاده از (۱-۱۲) داریم

$$\begin{aligned} H(Y) &= \int_0^{+\infty} -\ln(f_Y(y)) \cdot f_Y(y) dy = \int_0^{+\infty} -\ln \left(\frac{1}{\sqrt{y}} f_X(\sqrt{y}) \right) \cdot f_Y(y) dy \\ &= \int_0^{+\infty} \ln(\sqrt{y}) \cdot f_Y(y) dy - \int_0^{+\infty} \ln(f_X(\sqrt{y})) \cdot f_Y(y) dy \\ &= \int_0^{+\infty} \frac{1}{\sqrt{y}} \ln(\sqrt{y}) \cdot f_X(\sqrt{y}) dy - \int_0^{+\infty} \frac{1}{\sqrt{y}} \ln(f_X(\sqrt{y})) \cdot f_X(\sqrt{y}) dy. \end{aligned}$$

با تغییر متغیر $\sqrt{y} = u$ و جایگزینی، داریم

$$H(Y) = 2 \int_0^{+\infty} \ln(u) f_X(u) du - 2 \int_0^{+\infty} \ln(f_X(u)) f_X(u) du$$

با توجه به مقارن بودن تابع چگالی $f_X(x)$ ، نتیجه می‌شود

$$H(Y) = 2 \int_0^{+\infty} \ln(u) f_X(u) dy + H(X)$$

و برهان کامل می‌شود. $\square$

نتیجه ۳-۱-۴ ()** اگر $X \sim t_{(r)}$ باشد، در این صورت برای $Y = X^2$ داریم

$$H(Y) = \ln r + \ln B\left(\frac{1}{2}, \frac{r}{2}\right) + \left(\frac{r+1}{2}\right) \psi\left(\frac{r+1}{2}\right) + \frac{1}{2} \psi\left(\frac{1}{2}\right) - \frac{(r+2)}{2} \psi\left(\frac{r}{2}\right). \quad (۸-۳)$$

برهان. از آنجایی که Y دارای توزیع F با ۱ و r درجه آزادی است، با جایگزینی (۳-۶) در (۳-۷) به راحتی اثبات می‌شود.

۳-۲- اطلاع شانون مقادیر و زمان رکوردها برای توزیع t -استودنت بر اساس یک نمونه n تایی در مدل کلاسیک

در این قسمت، با استفاده از نتایج ارایه شده در فصل دوم، اطلاع شانون توزیع t -استودنت در مدل کلاسیک را مورد بررسی قرار می‌دهیم و مقادیر $(H_M^L(n))H_M^U(n)$ و $(H_{RT}^L(n))H_{RT}^U(n)$ را به ترتیب برای توزیع t -استودنت محاسبه نموده و در انتها به نتیجه‌گیری در مورد این مبحث می‌پردازیم. در ابتدا، یادآور می‌شویم که به دلیل خاصیت تقارن توزیع t -استودنت، فقط مقادیر $H_M^U(n)$ و $H_{RT}^U(n)$ را محاسبه می‌کنیم. برای محاسبه $H_M^U(n)$ با استفاده از (۲-۱) و (۲-۲) داریم

$$H_M^U(n) = - \int_{-\infty}^{+\infty} f_n(x) \ln f_n(x) dx$$

$$= -\ln n + \frac{n-1}{n} - n\phi_f^U(n-1)$$

$$= -\ln n + \frac{n-1}{n} - n \int_{-\infty}^{+\infty} (\ln f(x)) F^{n-1}(x) f(x) dx.$$

با جایگذاری (۳-۳) داریم

$$H_M^U(n) = -\ln n + \frac{n-1}{n} - n \int_{-\infty}^{+\infty} \ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\left(\frac{r+1}{2}\right)} \right) F^{n-1}(x) f(x) dx$$

$$H_M^U(n) = -\ln n + \frac{n-1}{n} - n \int_{-\infty}^{+\infty} \ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) F^{n-1}(x) f(x) dx$$

$$+ n \left(\frac{r+1}{2} \right) \int_{-\infty}^{+\infty} \ln \left(1 + \frac{x^2}{r} \right) F^{n-1}(x) f(x) dx.$$

با تغییر متغیر $F(x) = u$ در انتگرال اول، داریم

$$H_M^U(n) = -\ln n + \frac{n-1}{n} + \frac{1}{2} \ln r + \ln B\left(\frac{1}{2}, \frac{r}{2}\right)$$

$$+ n \left(\frac{r+1}{2} \right) \int_{-\infty}^{+\infty} \ln \left(1 + \frac{x^2}{r} \right) F^{n-1}(x) f(x) dx.$$

با توجه به غیرقابل محاسبه بودن انتگرال، جدول ۳-۱ مقدار $H_M^U(n)$ را برای درجه‌های آزادی

$r = 2, 3, 5, 10, 15, 20$ و حجم‌های نمونه $n = 2, 3, 5$ که با استفاده از نرم‌افزار *MAPLE* محاسبه

شده‌اند، نشان می‌دهد.

فصل سوم: اطلاع شانون مقادیر و زمان رکوردها در توزیع t -استودنت

جدول ۳-۱: $(H_M^U(n) = H_M^L(n))$ اطلاع شانون برای آخرین رکورد در مدل کلاسیک از توزیع t -استودنت

$r \backslash n$	۲	۳	۵	۱۰	۱۵	۲۰
۲	-۲/۱۵	-۱/۹۶	-۱/۸۲	-۱/۷۱	-۱/۶۷	۱/۶۶
۳	-۲/۶۴	-۲/۴۱	-۲/۲۴	-۲/۱۱	-۲/۰۷	-۲/۰۴
۵	-۳/۴۹	-۳/۱۹	-۲/۹۶	-۲/۷۹	-۲/۷۳	-۲/۷۰

همانطور که در جدول مشاهده می‌شود، با افزایش درجه آزادی اطلاع شانون نیز افزایش می‌یابد، در حالی که با افزایش n اطلاع شانون کم می‌شود.

اکنون می‌توانیم $H_{RT}^U(n)$ را با استفاده از قضیه ۲-۱-۱، رابطه (۲-۹) و نتیجه (۲-۱-۲) بدست آوریم. مقادیر آن برای درجه‌های آزادی $r = 2, 3, 5, 10, 15, 20$ و حجم‌های نمونه $n = 2, 3, 5$ در جدول ۳-۲ نشان داده شده‌اند.

جدول ۳-۲: $(H_{RT}^U(n) = H_{RT}^L(n))$ اطلاع شانون برای مقادیر و زمان رکوردها در مدل کلاسیک از توزیع t -استودنت

$r \backslash n$	۲	۳	۵	۱۰	۱۵	۲۰
۲	-۲,۶۹	-۲,۴۱	-۲,۱۹	-۲,۰۳	-۱,۹۷	-۱,۹۵
۳	-۳,۲۰	-۲,۸۴	-۲,۵۷	-۲,۳۶	-۲,۳۰	-۲,۲۷
۵	-۴,۰۰	-۳,۵۲	-۳,۱۵	-۲,۸۷	-۲,۷۸	-۲,۷۴

می‌توان مشاهده کرد با افزایش درجه آزادی، اطلاع شانون برای مقادیر و زمان رکوردها افزایش می‌یابد. بالعکس با افزایش n اطلاع شانون برای آن‌ها کم می‌شود.

۳-۳- اطلاع شانون مقادیر و زمان رکوردها در توزیع t -استودنت در طرح نمونه‌گیری معکوس

همان طور که در فصل دوم اشاره شد، در طرح نمونه‌گیری معکوس ابتدا تعداد رکوردها، m را مشخص می‌کنیم و نمونه‌گیری را تا رسیدن به m رکورد ادامه می‌دهیم. بنابراین اطلاع شانون آخرین رکورد بالایی (پایینی) و همه مقادیر رکورد بالایی (پایینی) از توزیع t -استودنت در طرح نمونه‌گیری معکوس را با استفاده از قضیه ۷-۱-۲ می‌توان بدست آورد. ابتدا به محاسبه اطلاع شانون آخرین رکورد بالایی می‌پردازیم، فرض کنید X دارای توزیع t -استودنت با r درجه آزادی باشد. با توجه به قضیه ۷-۱-۲ و رابطه (۱۲-۲) داریم

$$h_M^U(m) = \sum_{i=1}^{m-1} \ln i - (m-1)\psi(m) - \psi_f^U(m).$$

$\psi_f^U(m)$ را می‌توان با استفاده از (۱۵-۲) به صورت زیر بدست آورد.

$$\begin{aligned} \psi_f^U(m) &= \frac{1}{\Gamma(m)} \int_{-\infty}^{+\infty} (\ln f(x)) \times (-\ln(1-F(x)))^{m-1} f(x) dx \\ &= \frac{1}{\Gamma(m)} \int_{-\infty}^{+\infty} \ln \left(\frac{1}{\sqrt{r} \cdot B\left(\frac{1}{2}, \frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\left(\frac{r+1}{2}\right)} \right) \times (-\ln(1-F(x)))^{m-1} f(x) dx \\ &= \frac{1}{\Gamma(m)} \int_{-\infty}^{+\infty} \ln \left(\frac{1}{\sqrt{r} \cdot B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) \times (-\ln(1-F(x)))^{m-1} f(x) dx \\ &\quad - \frac{1}{\Gamma(m)} \cdot \left(\frac{r+1}{2}\right) \int_{-\infty}^{+\infty} \ln \left(1 + \frac{x^2}{r}\right) \times (-\ln(1-F(x)))^{m-1} f(x) dx. \end{aligned}$$

با استفاده از (۳-۳) داریم

$$\begin{aligned}\psi_f^U(m) &= \frac{1}{\Gamma(m)} \int_{-\infty}^{+\infty} \ln \left(\frac{1}{\sqrt{r} \cdot B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) \times (-\ln(1-F(x)))^{m-1} f(x) dx \\ &\quad - \frac{1}{\Gamma(m)} \cdot \left(\frac{r+1}{2}\right) \int_{-\infty}^{+\infty} \ln \left(1 + \frac{x^2}{r} \right) \times (-\ln(1-F(x)))^{m-1} f(x) dx \\ &= \frac{1}{\Gamma(m)} \frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \int_{-\infty}^{+\infty} \ln \left(\frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \right) \times (-\ln(1-F(x)))^{m-1} \left(1 + \frac{x^2}{r} \right)^{-\left(\frac{r+1}{2}\right)} dx \\ &\quad - \frac{1}{\Gamma(m)} \left(\frac{r+1}{2}\right) \frac{1}{\sqrt{r} B\left(\frac{1}{2}, \frac{r}{2}\right)} \int_{-\infty}^{+\infty} \ln \left(1 + \frac{x^2}{r} \right) \times (-\ln(1-F(x)))^{m-1} \left(1 + \frac{x^2}{r} \right)^{-\left(\frac{r+1}{2}\right)} dx.\end{aligned}$$

همان طور که مشاهده می‌شود به دلیل غیر قابل محاسبه بودن انتگرال، $\psi_f^U(m)$ را برای مقادیر

متفاوت $r=2,3,5,10$ و $m=1,2,3$ با استفاده از نرم افزار *Maple* محاسبه نموده که در جدول

۳-۳ ارائه شده است

جدول ۳-۳ $\psi_f^U(m)$

$m \backslash r$	۲	۳	۵	۱۰
۱	۱/۹۶	۱/۷۷	۱/۶۲	۱/۵۲
۲	۲/۴۹	۲/۲۲	۲/۰۱	۱/۸۶
۳	۳/۶۸	۳/۲۵	۲/۹۰	۲/۶۴

با استفاده از مقادیر جدول ۳-۳، $h_M^U(m)$ برای درجات آزادی $r = 2, 3, 5, 10$ و رکوردهای $m = 1, 2$ محاسبه شده و در جدول ۴-۳ ارائه شده است.

جدول ۴-۳: $(h_M^U(m) = h_M^L(m))$ اطلاع شانون برای آخرین رکورد در طرح نمونه‌گیری معکوس از توزیع t -استودنت

$m \backslash r$	۲	۳	۵	۱۰
۱	-۱/۹۶	-۱/۷۷	-۱/۶۲	-۱/۵۲
۲	-۲/۹۱	-۲/۶۴	-۲/۴۴	-۲/۲۸
۳	-۴/۸۴	-۴/۴۰	-۴/۰۵	-۳/۷۹

همانطور که در جدول مشاهده می‌شود، با افزایش درجه آزادی اطلاع شانون نیز افزایش می‌یابد، در حالی که با افزایش m اطلاع شانون کم می‌شود.

از مقادیر بدست آمده در جدول ۳-۳ و قضیه ۲-۱-۷ و رابطه (۲-۱۳) می‌توانیم $h_R^U(m)$ را برای درجات آزادی $r = 2, 3, 5, 10$ و رکوردهای $m = 1, 2, 3, 4, 5, 6, 7$ محاسبه کنیم، که این مقادیر در جدول ۵-۳ ارائه شده‌اند.

جدول ۳-۵: $(h_R^U(m) = h_R^L(m))$ اطلاع شانون برای مقادیر رکورد در طرح نمونه گیری معکوس از توزیع t -استودنت

$r \backslash m$	۲	۳	۵	۱۰
۱	-۱/۹۶	-۱/۷۷	-۱/۶۲	-۱/۵۲
۲	-۳/۴۵	-۳/۰۰	-۲/۶۴	-۲/۳۸
۳	-۵/۱۴	-۴/۲۵	-۳/۵۴	-۳/۰۳

همانطور که در جدول مشاهده می شود، با افزایش درجه آزادی اطلاع شانون نیز افزایش می یابد، در حالی که با افزایش n اطلاع شانون کم می شود.

با استفاده از مقادیر جدول ۳-۵ می توانیم $h_{RT}^U(m)$ را برای درجات آزادی $r = 2, 3, 5, 10$ و رکوردهای $m = 1, 2, 3$ محاسبه کنیم، که در جدول ۳-۶ ارائه شده اند.

جدول ۳-۶: $(h_{RT}^U(m) = h_{RT}^L(m))$ اطلاع شانون برای مقادیر و زمان رکوردها در طرح نمونه گیری معکوس از توزیع t -استودنت

$r \backslash m$	۲	۳	۵	۱۰
۱	-۱/۹۶	-۱/۷۷	-۱/۶۲	-۱/۵۲
۲	-۵/۰۹	-۴/۶۴	-۴/۲۹	-۴/۰۳
۳	-۹/۶۳	-۸/۷۴	-۸/۰۴	-۷/۵۲

۳-۴- نتیجه‌گیری و پیشنهادات

در این پایان‌نامه اطلاع شانون را برای مقادیر رکورد در نمونه‌هایی با اندازه ثابت (مدل کلاسیک) و طرح نمونه‌گیری معکوس برای توزیع t -استودنت مورد بررسی قرار دادیم. همان‌طور که از جداول ارایه شده در این فصل بر می‌آید، مقدار اطلاع شانون با افزایش درجه آزادی افزایش یافته و به توزیع نرمال استاندارد نزدیک می‌شود و با افزایش حجم نمونه، کاهش می‌یابد که بنابر تعریف آنتروپی مورد انتظار است.

پیشنهادهایی که برای آینده می‌توان در ادامه این پایان‌نامه مورد بررسی قرار داد، محاسبه اطلاع رنی برای مقادیر رکورد در نمونه‌هایی با اندازه ثابت (مدل کلاسیک) و طرح نمونه‌گیری معکوس برای توزیع t -استودنت و F می‌باشد.



پیوست

دستورهای *MAPLE*

دستور ۱. برای به دست آوردن اطلاع‌شان آخرین رکورد بالایی (پایینی) توزیع t -استودنت در مدل کلاسیک، جدول ۱-۳، از دستوره‌های زیر در نرم‌افزار MAPLE 14 استفاده می‌کنیم.

> restart

$$\begin{aligned} > f := (x, r) \rightarrow \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot \left(1 + \frac{x^2}{r}\right)^{-\frac{(r+1)}{2}} \\ f := (x, r) \rightarrow \frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right) \left(1 + \frac{x^2}{r}\right)^{-\frac{1}{2}r - \frac{1}{2}}}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)} \end{aligned}$$

$$\begin{aligned} > F(x, r) &:= \int_{-\infty}^x f(y, r) \, dy \\ F &:= (x, r) \rightarrow \int_{-\infty}^x f(y, r) \, dy \end{aligned}$$

$$\begin{aligned} > g(i, r) &:= \int_{-\infty}^{+\infty} \ln\left(1 + \frac{x^2}{r}\right) \cdot (F(x, r))^i \cdot f(x, r) \, dx \\ (i, r) &\rightarrow \int_{-\infty}^{\infty} \ln\left(1 + \frac{x^2}{r}\right) F(x, r)^i f(x, r) \, dx \end{aligned}$$

$$\begin{aligned} > G(i, r) &:= -\ln\left(\frac{\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \cdot \Gamma\left(\frac{r}{2}\right)}}{(i+1)}\right) + \left(\frac{(r+1)}{2} \cdot g(i, r)\right) \\ (i, r) &\rightarrow -\ln\left(\frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right)}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right) (i+1)}\right) + \left(\frac{1}{2}r + \frac{1}{2}\right) g(i, r) \end{aligned}$$

$$\begin{aligned} > H(n, r) &:= -\ln(n) + \frac{(n-1)}{n} - n \cdot G(n-1, r) \\ (n, r) &\rightarrow -\ln(n) + \frac{n-1}{n} - n G(n-1, r) \end{aligned}$$

دستور ۲. برای به دست آوردن شانون مقادیر و زمان‌های رکورد بالایی (پایینی) برای توزیع

t - استودنت در مدل کلاسیک، جدول ۳-۲، از دستوره‌های زیر در نرم افزار MAPLE 14 استفاده

می کنیم.

> restart

$$> f := (x, r) \rightarrow \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot \left(1 + \frac{x^2}{r}\right)^{-\frac{(r+1)}{2}}$$

$$f := (x, r) \rightarrow \frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right) \left(1 + \frac{x^2}{r}\right)^{-\frac{1}{2}r - \frac{1}{2}}}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)}$$

$$> F(x, r) := \int_{-\infty}^x f(y, r) \, dy$$

$$F := (x, r) \rightarrow \int_{-\infty}^x f(y, r) \, dy$$

$$> g(i, r) := \int_{-\infty}^{+\infty} \ln\left(1 + \frac{x^2}{r}\right) \cdot (F(x, r)^i) \cdot f(x, r) \, dx$$

$$(i, r) \rightarrow \int_{-\infty}^{\infty} \ln\left(1 + \frac{x^2}{r}\right) F(x, r)^i f(x, r) \, dx$$

$$> G(i, r) := -\ln\left(\frac{\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \cdot \Gamma\left(\frac{r}{2}\right)}}{(i+1)}\right) + \left(\frac{(r+1)}{2} \cdot g(i, r)\right)$$

$$(i, r) \rightarrow -\ln\left(\frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right)}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right) (i+1)}\right) + \left(\frac{1}{2}r + \frac{1}{2}\right) g(i, r)$$

$$> H(n, r) := \Psi(n+1) + \gamma - \sum_{i=1}^n \frac{1}{(i)^2} - \sum_{i=0}^{n-1} G(i, r)$$

$$(n, r) \rightarrow \Psi(n+1) + \gamma - \left(\sum_{i=1}^n \frac{1}{i^2}\right) - \left(\sum_{i=0}^{n-1} G(i, r)\right)$$

دستور ۳. برای به‌دست‌آوردن اطلاع‌شانون آخرین رکورد بالایی (پایینی) برای توزیع t - استودنت

در طرح نمونه‌گیری معکوس جدول ۳-۴، از دستوره‌های زیر در نرم افزار MAPLE 14 استفاده

می‌کنیم.

> restart

$$> f := (x, r) \rightarrow \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \Gamma\left(\frac{r}{2}\right)} \cdot \left(1 + \frac{x^2}{r}\right)^{-\frac{(r+1)}{2}}$$

$$f := (x, r) \rightarrow \frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right) \left(1 + \frac{x^2}{r}\right)^{-\frac{1}{2}r - \frac{1}{2}}}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)}$$

$$> F(x, r) := \int_{-\infty}^x f(y, r) \, dy$$

$$F := (x, r) \rightarrow \int_{-\infty}^x f(y, r) \, dy$$

$$> g(i, r) := \frac{1}{\Gamma(i)} \cdot \int_{-\infty}^{+\infty} \ln\left(1 + \frac{x^2}{r}\right) \cdot \left(-\ln(F(x, r))\right)^{i-1} \cdot f(x, r) \, dx$$

$$(i, r) \rightarrow \int_{-\infty}^{\infty} \ln\left(1 + \frac{x^2}{r}\right) F(x, r)^i f(x, r) \, dx$$

$$> G(i, r) := -\ln\left(\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \cdot \Gamma\left(\frac{r}{2}\right)}\right) + \left(\frac{r+1}{2} \cdot g(i, r)\right)$$

$$(i, r) \rightarrow -\ln\left(\frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right)}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)}\right) + \left(\frac{1}{2}r + \frac{1}{2}\right) g(i, r)$$

$$> h(m, r) := \sum_{i=1}^{m-1} \ln(i) - (m-1)\Psi(m) - G(i, r)$$

$$h := (m, r) \rightarrow \sum_{i=1}^{m-1} \ln(i) - (m-1)\Psi(m) - G(i, r)$$

> evalf(h(m, r));

$$\ln(\Gamma(m)) - 1. (m-1.) \Psi(m) - 1. G(i, r)$$

دستور ۴. برای به دست آوردن اطلاع شانون مقادیر رکورد بالایی (پایینی) برای توزیع t -استودنت در طرح نمونه‌گیری معکوس جدول ۳-۵، از دستوره‌های زیر در نرم افزار MAPLE 14 استفاده می‌کنیم.

> restart

$$\begin{aligned} > f := (x, r) \rightarrow \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \Gamma\left(\frac{r}{2}\right)} \cdot \left(1 + \frac{x^2}{r}\right)^{-\frac{(r+1)}{2}} \\ f := (x, r) \rightarrow \frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right) \left(1 + \frac{x^2}{r}\right)^{-\frac{1}{2}r - \frac{1}{2}}}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)} \end{aligned}$$

$$> F(x, r) := \int_{-\infty}^x f(y, r) \, dy$$

$$F := (x, r) \rightarrow \int_{-\infty}^x f(y, r) \, dy$$

$$> g(i, r) := \frac{1}{\Gamma(i)} \cdot \int_{-\infty}^{+\infty} \ln\left(1 + \frac{x^2}{r}\right) \cdot \left(-\ln(F(x, r))\right)^{i-1} \cdot f(x, r) \, dx$$

$$(i, r) \rightarrow \int_{-\infty}^{\infty} \ln\left(1 + \frac{x^2}{r}\right) F(x, r)^i f(x, r) \, dx$$

$$> G(i, r) := -\ln\left(\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \cdot \Gamma\left(\frac{r}{2}\right)}\right) + \left(\frac{(r+1)}{2} \cdot g(i, r)\right)$$

$$(i, r) \rightarrow -\ln\left(\frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right)}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)}\right) + \left(\frac{1}{2}r + \frac{1}{2}\right) g(i, r)$$

$$> h(m) = -\sum_{i=1}^m \psi(i) + \frac{m \cdot (m-1)}{2}$$

$$h(R) = -\left(\sum_{i=1}^m \psi(i)\right) + \frac{1}{2} m (m-1)$$

> evalf(h(m))

$$h(m)$$

دستور ۵. برای به دست آوردن اطلاع شانون مقادیر و زمان‌های رکورد بالایی (پایینی) برای توزیع

t -استودنت در طرح نمونه‌گیری معکوس جدول ۳-۵، از دستورهایی زیر در نرم‌افزار MAPLE 14

استفاده می‌کنیم.

> restart

$$> f := (x, r) \rightarrow \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \Gamma\left(\frac{r}{2}\right)} \cdot \left(1 + \frac{x^2}{r}\right)^{-\frac{(r+1)}{2}}$$

$$f := (x, r) \rightarrow \frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right) \left(1 + \frac{x^2}{r}\right)^{-\frac{1}{2}r - \frac{1}{2}}}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)}$$

$$> F(x, r) := \int_{-\infty}^x f(y, r) \, dy$$

$$F := (x, r) \rightarrow \int_{-\infty}^x f(y, r) \, dy$$

$$> g(i, r) := \frac{1}{\Gamma(i)} \cdot \int_{-\infty}^{+\infty} \ln\left(1 + \frac{x^2}{r}\right) \cdot \left(-\ln(F(x, r))\right)^{i-1} \cdot f(x, r) \, dx$$

$$(i, r) \rightarrow \int_{-\infty}^{\infty} \ln\left(1 + \frac{x^2}{r}\right) F(x, r)^i f(x, r) \, dx$$

$$> G(i, r) := -\ln\left(\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi \cdot r} \cdot \Gamma\left(\frac{r}{2}\right)}\right) + \left(\frac{(r+1)}{2} \cdot g(i, r)\right)$$

$$(i, r) \rightarrow -\ln\left(\frac{\Gamma\left(\frac{1}{2}r + \frac{1}{2}\right)}{\sqrt{\pi r} \Gamma\left(\frac{1}{2}r\right)}\right) + \left(\frac{1}{2}r + \frac{1}{2}\right) g(i, r)$$

$$> h(m) := hR(m) - \sum_{i=2}^m (m-i+1) \zeta(i);$$

$$h := m \rightarrow hR(m) - \left(\sum_{i=2}^m (m-i+1) \zeta(i)\right)$$

$$> evalf(h(m))$$

$$h(m)$$

دستور ۶. دستور رسم شکل (۳-۱)، تابع چگالی احتمال توزیع t -استودنت در MAPLE 14.

> *with(plots) :*
 > *setoptions(title = 'Family Plot', axes = boxed)*
 >
$$\text{plot} \left(\left[\left[\text{seq} \left(\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r \cdot \pi} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}}, r = 1..4 \right), x = -4.. \right. \right. \right. \\ \left. \left. \left. + 4 \right] \right] \right)$$

دستور ۷. دستور رسم شکل (۳-۲)، تابع توزیع احتمال t -استودنت، در MAPLE 14.

> *with(plots) :*
 > *setoptions(title = 'Family Plot', axes = boxed)*
 >
$$\text{plot} \left(\left[\left[\text{seq} \left(\frac{1}{2} \right. \right. \right. \right. \\ \left. \left. \left. + \frac{x \cdot \Gamma\left(\frac{1}{2} r + \frac{1}{2}\right) \text{hypergeom}\left(\left[\frac{1}{2}, \frac{1}{2} r + \frac{1}{2}\right], \left[\frac{3}{2}\right], -\frac{x^2}{r}\right)\right]}{\sqrt{\pi r} \Gamma\left(\frac{1}{2} r\right)}, 1 \right. \right. \right. \right. \\ \left. \left. \left. = 1..4 \right] \right] \right), x = -4..4 \right)$$

منابع و مراجع

- [۱] آرشی، محمد، ۱۳۸۴، استنباط آماری بر اساس مقادیر رکوردی و زمان رخداد رکوردها، پایان نامه کارشناسی ارشد، دانشگاه فردوسی مشهد.
- [2] Madadi, M., **(2009)**, Ph.D. thesis, "Entropy in record data", Math. & Computer Faculty, Bamonar University
- [3] Arnold, B. C., Balakrishnan, N., Nagaraja, H. N., **(1998)**, "**Records**", John Wiley Sons, New York
- [4] Tata, M. N., **(1969)**, "On outstanding values in a sequence of variables" , Zeitschrift fuer Wahrscheinlichkeitstheorie und Verwandte Gebiete, Vol **12**, P.P **9-20**
- [5] Galambos, **(1987)**, "The Asymptotic Theory of Extreme Order Statistics" , Krieger, Malabar, Florida
- [6] Samaniego, F.J. Whitaker, L. R., **(1986)**, "On estimating population characteristics from record-breaking observation", I: Parametric results, Naval Research Logistics Quarterly, Vol **33**, P.P **531-543**
- [7] Ahmadi, J., Arghami, N. R., **(2003)**, "Nonparametric confidence and tolerance intervals based on record data", Statistical Papers, Vol **44**, P.P **455-468**
- [8] Ahmed, N. A., Gokhale, D. V., **(1989)**, "Entropy expressions and their estimators for multivariate distributions" IEEE Transactions on information Theory, Vol **35**, P.P **688-692**
- [9] Gulati, S., Padgett, W. J., **(2003)**, "Parametric and Nonparametric Inference from Record-breaking Data", Lecture Notes in Statistics, Springer Verlag, New York
- [10] Ahmadi, J., Arghami, N. R., **(2001)**, "On the Fisher information in record values", Metrika, Vol **53**, P.P **195-206**
- [11] Darbellay, G. A., Vajda, I., **(2000)**, "Entropy expressions for multivariate continuous distributions", IEEE Transactions on Information Theory, Vol **46**, P.P **709-712**

- [12] Tsallis, C., **(1988)**, "Possible generalization of Boltzmann-Gibbs statistics", Journal of Statistical Physics, Vol 52, P.P 479-487
- [13] Nyquist, H., **(1924)**, "Certain affecting telegraph speed", Bell System Technical Journal, Vol 3, P.P 324-346
- [14] Nyquist, H., **(1928)**, "Certain topics in telegraph transmission theory", AIEEE Transactions, Vol 47, P.P 617-644
- [15] Hartley, R. T. V., **(1928)**, "Tranmission of information", Bell System Technical Journal, Vol 7, P.P 535-563
- [16] Hofmann, G., Nagaraja, H. N., **(2003)**, "Fisher information in record data", Metrika, Vol 57, P.P 177-193
- [17] Cover, T. M., Thomas, J. A., **(2006)**, "Elements of Information Theory", John Wiley, New York
- [18] Harris, B., **(1982)**, "Entropy", in: S. Kotz, N. L. Johanson (Eds.), Encyclopedia of Statistical Sciences, Vol 2, P.P 512-516
- [19] Nayak, T. K., **(1985)**, "On diversity measures based on entropy functions", Communication in Statistics Theory and Methods, Vol 14, P.P 203-215
- [20] Soofi, E. S., **(2000)**, "Principal information theoretic approaches", Journal of the American Statistical Association, Vol 95, P.P 1349-1353
- [21] Dicrescenzo, A., Longobardi, M., **(2002)**, "Entropy-based measure of uncertainty in past lifetime distributions", Journal of Applied Probability, Vol 39, P.P 434-440
- [22] Dicrescenzo, A., Longobardi, M., **(2004)**, "A measure of discrimination between past life-time distributions", Statistics and Probability Letters, Vol 67, P.P 173-182

- [23] Ebrahimi, N., **(1996)**, " How to measure uncertainty in the residual lifetime distribution", Sankhya A, Vol **58**, P.P **48-57**
- [24] Ebrahimi, N., Kirmani S. N. U. A., **(1996)**, "A characterization of the proportional hazards model through a measure of discrimination between two residual life distributions", Biometrika, Vol **83**, P.P **233-235**
- [25] Ebrahimi,N., Kirmani S. N. U. A., **(1996)**, "Some results on ordering of survival functions through uncertainty", Statistics and Probability Letters, Vol **29**, P.P **167-176**
- [26] Baratpour, S., Ahmadi, J., Arghami, N. R., **(2007)**, "Entropy properties of record statistics", Statistical Papers, Vol **48**, P.P **197-213**
- [27] Zahedi, H., Shakil, M., **(2006)**, "Properties of entropies of record values in reliability and life testing context", Communication in Statistics Theory and Methods, Vol **35**, P.P **977-1010**
- [28] Lazo, A. C. G., Rathie, P. N., **(1978)**, "On the entropy of continuous distribution", IEEE Transactions on information Theory, Vol **24**, P.P **120-122**
- [29] R'enyi, A, (1961), "On Measures of entropy and information", Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability, University of California Press, Vol **1**, P.P **547 -561**
- [30] Tata, M. N., Madadi, M., (2011), " Shanon information in record data", Metrika, Vol **74**, P.P **11-31**
- [31] Nadarjah , S., Zografos, K., **(2003)** "Formulas for Renyi information and related measures for univariate distributions", Information Sciences, Vol **155**, P.P **119-138**

Abstract

Shanon entropy (shanon information) has been recently investigated by several authors such as Baratpor and Zahedi, This entropy can be successfully employed to quantify the extent of information (regarding the family of distributions). Madady and Tata (2011), for a sequence of continuous random variables (iid), have derived some relations for Shanon information of maximum record and the values and times of records occurrence in a classical model (a random sample with a size of n) and inverse sampling plan. This information can be calculated for well-known statistical distributions. In the thesis Shanon information of upper and lower record values and records occurrence times are investigated for the well-known t-student distribution. This thesis includes 3 chapters. In chapter 1, the history of records and entropy and some definitions and the applications of record values are presented. In chapter 2, the results for Shanon information of upper and lower record values and the relation between records occurrence time in classical model and inverse sampling plan are discussed. In chapter 3, Shanon information of records occurrence time and values are calculated for t-student distribution. In appendix A, the programs and commands written with Maple software are presented. In this thesis, the cases where the propositions are already available, but theorems are by the author, have indicated by (*) and the cases in which both the proposition and the theorem are new, have indicated by (**).