



استنباط آماری ۲

دکتر پرویز نصیری

امروز کتابخوانی و علم‌آموزی نه تنها یک وظیفه‌ی ملی، که یک واجب دینی است^۱.

مقام معظم رهبری

در عصر حاضر یکی از شاخصه‌های ارزیابی رشد، توسعه و پیشرفت فرهنگی هر کشوری میزان تولید کتاب، مطالعه و کتابخوانی مردم آن مرز و بوم است. ایران اسلامی نیز از دیرباز تاکنون با داشتن تمدنی چندهزارساله و مراکز متعدد علمی، فرهنگی، کتابخانه‌های معتبر، علما و دانشمندان بزرگ با آثار ارزشمند تاریخی، سرآمد دولت‌ها و ملت‌های دیگر بوده و در عرصه فرهنگ و تمدن جهانی به‌سان خورشیدی تابناک همچنان می‌درخشد و با فرزندان نیک‌نهاد خویش هنرنمایی می‌کند. چه کسی است که در دنیا با دانشمندان فرزانه و نام‌آور ایرانی همچون ابوعلی سینا، ابوریحان بیرونی، فارابی، خوارزمی و ... همچنین شاعران برجسته‌ای نظیر فردوسی، سعدی، مولوی، حافظ و ... آشنا نباشد و در مقابل عظمت آنها سر تعظیم فرود نیاورد. تمامی این افتخارات ارزشمند، برگرفته از میزان عشق و علاقه فراوان ملت ما به فراگیری علم و دانش از طریق خواندن و مطالعه منابع و کتاب‌های گوناگون است. به شکرانه الهی، تاریخ و گذشته ما، همیشه درخشان و پربار است. ولی اکنون در این زمینه در چه جایگاهی قرار داریم؟ آمار و ارقام ارائه‌شده از سوی مجامع و سازمان‌های فرهنگی در مورد سرانه مطالعه هر ایرانی، برایمان چندان امیدوارکننده نمی‌باشد.

کتاب، دروازه‌ای به سوی گستره دانش و معرفت است و کتاب خوب، یکی از بهترین ابزارهای کمال بشری است. همه دستاوردهای بشر در سراسر عمر جهان، تا آنجا که قابل کتابت بوده است، در میان دست‌نوشته‌هایی است که انسان‌ها پدید آورده و می‌آورند. در این مجموعه بی‌نظیر، تعالیم الهی، درس‌های پیامبران به بشر، و همچنین علوم مختلفی است که سعادت بشر بدون آگاهی از آنها امکان‌پذیر نیست. کسی که با دنیای زیبا و زندگی‌بخش کتاب ارتباط ندارد بی‌شک از مهم‌ترین دستاورد انسانی و نیز از بیشترین معارف الهی و بشری محروم است. با این دیدگاه، به‌روشنی می‌توان ارزش و مفهوم رمزی عمیق در این حقیقت تاریخی را دریافت که اولین خطاب خداوند متعال به پیامبر گرامی اسلام(ص) این است که «بخوان!» و در اولین سوره‌ای که بر آن فرستاده عظیم‌الشان خداوند، فرود آمده، نام «قلم» به تجلیل یاد

1. <https://farsi.khamenei.ir/message-content?id=2696>

شده است: «إِقْرَأْ وَ رَبُّكَ الْأَكْرَمُ. الَّذِي عَلَّمَ بِالْقَلَمِ» در اهمیت عنصر کتاب برای تکامل جامعه انسانی، همین بس که تمامی ادیان آسمانی و رجال بزرگ تاریخ بشری، از طریق کتاب جاودانه مانده اند.

دانشگاه پیام نور با گستره جغرافیایی ایران شمول خود با هدف آموزش برای همه، همه جا و همه وقت، به عنوان دانشگاهی کتاب محور در نظام آموزش عالی کشورمان، افتخار دارد جایگاه اندیشه سازی و خردورزی بخش عظیمی از جوانان جویای علم این مرز و بوم باشد. تلاش فراوانی در ایام طولانی فعالیت این دانشگاه انجام پذیرفته تا با بهره گیری از تجربه های گرانقدر استادان و صاحب نظران برجسته کشورمان، کتاب ها و منابع آموزشی درسی شاخص و خودآموز تولید شود. در آینده هم، این مهم با هدف ارتقای سطح علمی، روزآمدی و توجه بیشتر به نیازهای مخاطبان دانشگاه پیام نور با جدیت ادامه خواهد داشت. به طور قطع استفاده از نظرات استادان، صاحب نظران و دانشجویان محترم، ما را در انجام این وظیفه مهم و خطیر یاری رسان خواهد بود. پیشاپیش از تمامی عزیزانی که با نقد، تصحیح و پیشنهادهای خود ما را در انجام این وظیفه خطیر یاری می رسانند، سپاسگزاری می نمایم. لازم است از تمامی اندیشمندانی که تاکنون دانشگاه پیام نور را منزلگاه اندیشه سازی خود دانسته و ما را در تولید کتاب و محتوای آموزشی درسی یاری نموده اند، صمیمانه قدردانی گردد. موفقیت و بهروزی تمامی دانشجویان و دانش پژوهان عزیز آرزوی همیشگی ما است.

دانشگاه پیام نور

فهرست مطالب

فصل اول.

فصل اول. کلیات

۱	مقدمه
۱	۱-۱ آماره
۲	۲-۱ آماره بسنده
۸	۳-۱ کامل بودن
۱۰	۴-۱ کمیت محوری
۱۱	۵-۱ خانواده نمایی یک پارامتری
۱۴	۶-۱ خانواده نمایی دو یا چند پارامتری
۱۸	۷-۱ خانواده پیتمن
۱۹	۸-۱ نسبت درستمایی یکنوا
۲۱	۹-۱ برآورد ماکسیمم درستمایی در خانواده نمایی
۲۴	۱۰-۱ نامساوی کرامر- رائو
۲۷	۱۱-۱ خانواده کرامر
۲۹	خودآزمایی

فصل دوم. برآوردگرهای سازگار

۳۵	مقدمه
۳۵	۱-۲ برآوردگرهای سازگار

۴۰	۲-۲ مولد برآوردگرهای سازگار
۴۵	۳-۲ بردار برآوردگرهای سازگار
۴۷	۴-۲ برآوردگرهای نرمال مجانبی
۵۰	۵-۲ توزیع مجانبی آماره مرتبه $1-\alpha$
۵۲	۶-۲ توزیع تابعی از برآوردگر نرمال مجانبی
۵۸	خودآزمایی

فصل سوم. آزمون فرضیه‌های آماری

۶۳	مقدمه
۷۱	۱-۳ پرتوان‌ترین آزمون
۷۳	۲-۳ لم نیمن پیرسن
۸۴	۳-۳ پرتوان‌ترین آزمون در خانواده نمایی
۹۶	۴-۳ پرتوان‌ترین آزمون به‌طور یکنواخت
۹۸	۵-۳ پرتوان‌ترین آزمون‌های یکنواخت خانواده نمایی
۱۰۵	۶-۳ خواص نسبت درستنمایی یکنوا
۱۱۲	۷-۳ آزمون‌های اجتماع-اشتراک و اشتراک-اجتماع
۱۱۲	۱-۷-۳ روش آزمون اجتماع-اشتراک
۱۱۵	۲-۷-۳ روش آزمون اشتراک-اجتماع
۱۱۷	۸-۳ پرتوان‌ترین آزمون یکنواخت برای فرض‌های دوطرفه
۱۲۱	خودآزمایی

فصل چهارم. آزمون نسبت درستنمایی تعمیم‌یافته

۱۲۹	مقدمه
۱۳۸	۱-۴ آزمون‌های k نمونه‌ای
۱۴۳	۲-۴ آزمون‌های بزرگ نمونه‌ای
۱۴۷	۳-۴ آزمون‌های شرطی
۱۵۰	۴-۴ آزمون فرض در جداول توافقی
۱۵۲	۵-۴ آزمون استقلال دو متغیر
۱۵۴	۶-۴ آزمون فرض با تقریب توزیع پسین در استنباط بیزی
۱۵۶	۷-۴ تقریب نرمال توزیع پسین

۱۶۳	خودآزمایی
۱۶۹	فصل پنجم. برآورد فاصله‌ای
۱۶۹	مقدمه
۱۷۱	۵-۱ فاصله اطمینان دوطرفه
۱۷۳	۵-۲ فواصل اطمینان یک‌طرفه
۱۷۵	۵-۳ روش کمیت محوری
۱۸۷	۵-۴ فاصله اطمینان برای تفاضل میانگین‌های دو جامعه نرمال مستقل
۱۸۹	۵-۵ فاصله اطمینان با استفاده از CDF
۱۹۵	۵-۶ یافتن فاصله اطمینان به روش آماری
۲۰۰	۵-۷ کوتاه‌ترین فواصل اطمینان
۲۰۴	۵-۸ برآورد فاصله‌ای بیزی
۲۱۵	۵-۹ فاصله اطمینان HPD
۲۱۸	خودآزمایی
۲۲۵	فصل ششم. فواصل اطمینان مجانبی
۲۲۵	مقدمه
۲۲۶	۶-۱ فاصله اطمینان برای $\tau(\theta)$ وقتی که $I_x(\theta)$ تابعی از θ نیست
۲۲۹	۶-۲ فاصله اطمینان برای $\tau(\theta)$ وقتی که $I_x(\theta)$ تابعی از θ است
۲۳۹	۶-۳ فاصله اطمینان مجانبی با استفاده از تابع امتیاز
۲۴۳	۶-۴ فاصله اطمینان مجانبی برای $\tau(\theta)$
۲۴۸	خودآزمایی
۲۵۱	منابع
۲۵۳	پیوست‌ها
۲۵۳	۱- توزیع‌های مهم آماری
۲۵۷	۲- جداول توزیع‌های پیشین مزدوج
۲۵۸	۳- سؤالات تشریحی
۲۶۵	۴- سؤالات تستی
۳۰۱	۵- جداول آماری

پیشگفتار

کتاب حاضر مقدمه‌ای بر استنباط آماری است و متناسب با سرفصل وزارت علوم، تحقیقات و فناوری برای درس استنباط آماری II دانشجویان مقطع کارشناسی ارشد رشته آمار دانشگاه پیام‌نور نگارش شده است، که برای سایر رشته‌های علوم پایه و فنی مهندسی که مباحث مرتبط دارند قابل استفاده است. برای نگارش آن از منابع مختلف، علی‌الخصوص مبانی آمار ریاضی آقای دکتر احمد پارسیان و سایر منابع مرتبط با استنباط آماری استفاده شده است که در فهرست منابع ذکر شده‌اند. مطالب در شش فصل به شرح ذیل آورده شده است که:

- در فصل اول آماره، آماره بسنده و بسنده مینیمال، کامل بودن، کمیت محوری، خانواده توزیع نمایی یک، دو و چندپارامتری، خانواده پیتمن، خانواده کرامر، نسبت درستنمایی یکنوا و برآوردگر ماکسیمم درستنمایی به همراه نامساوی کرامر- رائو آورده شده است.

- در فصل دوم پس از معرفی برآوردهای سازگار، مولد برآوردهای سازگار، برآوردهای نرمال مجانبی به همراه تابعی از برآوردهای نرمال مجانبی آورده شده است.

- در فصل سوم پس از طرح مسئله آزمون فرض‌ها، لم‌نیمن- پیرسن ارائه شده است و با استفاده از تکنیک لم‌نیمن- پیرسن پرتوان‌ترین آزمون و پرتوان‌ترین آزمون یکنواخت فرض‌های ساده یا مرکب در مقابل فرض‌های ساده یا مرکب مورد بحث قرار گرفته‌اند و در انتها پس ارائه آزمون‌های اجتماع- اشتراک و آزمون‌های اشتراک- اجتماع، پرتوان‌ترین آزمون یکنواخت برای فرض‌های ساده در مقابل فرض‌های دوطرفه مورد بحث قرار می‌گیرد.

- در فصل چهارم آزمون نسبت درستمایی تعمیم‌یافته ارائه گردیده و آزمون‌های k نمونه‌ای، آزمون‌های بزرگ نمونه‌ای و آزمون‌های شرطی، آزمون فرض در جداول توافقی، آزمون فرض با تقریب توزیع پسین بیزی و تقریب نرمال توزیع پسین جداگانه مورد بحث قرار گرفته‌اند.

- در فصل پنجم پس از طرح برآورد فاصله‌ای، برآورد فاصله‌ای برای پارامتر یا تابعی از آن به روش‌های کمیت محوری، CDF، آماری، بیزی، HPD و کوتاه‌ترین فواصل اطمینان ارائه شده است.

- در فصل ششم فاصله اطمینان مجانبی برای تابعی از پارامتر با توجه به رفتار تابع اطلاع فیشر، استفاده از تابع امتیاز و فاصله اطمینان مجانبی برای $\tau(\theta)$ ارائه شده است.

با توجه به سابقه تدریس بعضی از فصل‌ها در مقایسه با سرفصل موجود برای استفاده دانشجویان محترم در تحقیقات توسعه داده شده است. در اینجا لازم می‌دانم از دفتر تدوین دانشگاه پیام نور، همکاران علمی علی‌الخصوص اعضای محترم شورای تخصصی گروه آمار، آقای دکتر یوسفی، سرکار خانم دکتر عباسی و آقای دکتر رضایی رکن‌آبادی عضو محترم هیئت علمی دانشگاه فردوسی مشهد که در مراحل داوری، ویرایش و داوری نهایی همکاری علمی در چاپ اول با اینجانب داشته‌اند تشکر کنم. با توجه به استقبال دانشگاه‌های مختلف از کتاب به

عنوان منبع برای استنباط آماری ۲ و ارسال نکات ویرایشی و تکمیلی از بعضی دانشگاه‌ها فرصت حاصل گردید که موارد دریافتی بررسی و اعمال گردد. یکی از اولین کسانی که نکات علمی و ویرایشی را در ماه‌های اول پس از چاپ اول ارسال کردند جناب آقای دکتر مسعود گنجی استاد تمام گروه آمار دانشگاه محقق اردبیلی بود که تمام تذکرات با کمال افتخار در چاپ دوم اعمال گردید. همچنین از زحمات جناب آقای دکتر غلامرضا حسامیان که داوری چاپ دوم را تقبل و نکات علمی و ویرایش برای ارتقاء کتاب داشته‌اند کمال تشکر را دارم. در پایان از سرکار خانم فاطمه محمدی که حروفچینی و صفحه‌آرایی نسخه چاپ اول و دوم کتاب را بر عهده داشتند تشکر می‌نمایم.

بدیهی است که در تدوین هر اثر علمی لغزش و خطا اجتناب‌ناپذیر است از این رو انتظار می‌رود اساتید محترم و دانشجویان ارجمند نظرات اصلاحی خود را تذکر دهند.

پرویز نصیری

استاد تمام و عضو هیئت علمی دانشگاه پیام نور

پاییز ۱۴۰۳

فصل اول

کلیات

مقدمه

در این فصل برای یادآوری مباحث آماره، آماره بسنده، آماره بسنده مینیمال، کامل بودن، کمیت محوری، خانواده نمایی یک، دو و چندپارامتری، خانواده پیتمن، خانواده کرامر، نسبت درستنمایی یکنوا، برآورد ماکسیمم درستنمایی و نامساوی کرامر- رائو را با ارائه مثال‌های مورد بحث قرار می‌گیرد. آشنایی با مباحث فوق کمک شایانی برای درک مطالب ارائه شده در فصل‌های بعدی خواهد بود.

۱-۱ آماره^۱

فرض کنید نمونه تصادفی $\mathbf{x} = (X_1, X_2, \dots, X_n)$ از مدل مورد بررسی دارای فضای پارامتری Θ باشد، به طوری که بتوان فضای پارامتر را به صورت $\Theta = \{\theta | \theta \in B\}$ نوشت. که در آن B زیر مجموعه‌ای از اعداد حقیقی است. چون مقدار θ مجهول است، علاقه‌مند به کسب اطلاعات با استفاده از نمونه تصادفی $\mathbf{x}$ درباره پارامتر θ یا تابعی از آن ($\tau(\theta)$) هستیم. برای این منظور آماره T یا $T(\mathbf{x})$ که تابعی از $\mathbf{x}$ است در نظر گرفته می‌شود، به طوری که برای مقدار مشاهده $\mathbf{x}, \mathbf{x} = x$ مقدار $T(\mathbf{x})$ مقداری ثابت است. برای مثال اگر $\mathbf{x} = (X_1, X_2, \dots, X_n)$ نمونه تصادفی از یک مدل یا توزیع آماری مورد بررسی باشد. توابع

$$T_1(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n X_i$$

$$T_r(\mathbf{x}) = T(X_{(1)}, X_{(r)}, \dots, X_{(n)}) = X_{(n)} - X_{(1)}, \quad X_{(1)} \leq X_{(r)} \leq \dots \leq X_{(n)}$$

$$T_r(\mathbf{x}) = \frac{(n-1)S^r}{\sigma^r}, \quad S^r = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^r \quad \text{و} \quad \sigma^r \text{ معلوم}$$

$$T_p = \frac{\bar{X}_1 - \bar{X}_r}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_r}}}, \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_r-1)S_r^2}{n_1 + n_r - 2}$$

آماره هستند. ضمناً هر آماره $T(\mathbf{x})$ یک متغیر تصادفی است و مدل یا توزیع مربوط به خود را دارد. که ممکن است توزیع آن به پارامتر θ بستگی داشته باشد و یا نه. ولی مهم این است که ممکن است آماره $T(\mathbf{x})$ اطلاعاتی درباره θ یا $\tau(\theta)$ داشته باشد، همان‌طور که نمونه تصادفی $\mathbf{x}$ دارد. سؤال اساسی این است که چگونه می‌توان برای هر نمونه تصادفی $\mathbf{x}$ ، آماره $T(\mathbf{x})$ را طوری تعیین کرد که اطلاعات یکسانی درباره پارامتر θ داشته باشد.

۲-۱ آماره بسنده

مفهوم آماره بسنده اولین بار در دهه ۱۹۲۰ توسط فیشر^۱ ارائه شد. او برای برآورد واریانس توزیع نرمال دو آماره $T_1(\mathbf{x}) = \sum_{i=1}^n |X_i - \bar{X}|^2$ و $T_r(\mathbf{x}) = \sum_{i=1}^n (X_i - \bar{X})^2$ را در نظر گرفت و نشان داد که توزیع $T_1(\mathbf{x})$ به شرط $T_r(\mathbf{x}) = t$ به پارامتر σ بستگی ندارد در حالی که توزیع $T_r(\mathbf{x})$ به شرط $T_1(\mathbf{x}) = t$ به σ بستگی دارد. بنابراین می‌توان نتیجه گرفت که همه اطلاعات درباره σ^2 در آماره $T_r(\mathbf{x})$ نهفته است. و هر برآوردی برای σ^2 باید تابعی از $T_r(\mathbf{x})$ باشد؛ یعنی اینکه هر برآورد $\hat{\sigma}^2$ براساس آماره $T_1(\mathbf{x})$ می‌تواند با استفاده از اطلاعات آماره $T_r(\mathbf{x})$ بهبود پیدا کند. در حالی که آماره

1. Fisher

۳ کلیات

$T_r(\mathbf{x})$ نمی‌تواند با استفاده از اطلاعات آماره $T_1(\mathbf{x})$ بهبود پیدا کند. برای تعمیم ایده فیشر، فرض کنید $\mathbf{x} = (X_1, X_2, \dots, X_n)$ یک نمونه تصادفی از تابع توزیع F_θ باشد. برای $\theta \in \Theta$ ، $A_t = \{\mathbf{x} : T(\mathbf{x}) = t\}$ تکیه‌گاه آماره $T(\mathbf{x})$ باشد. اگر توزیع شرطی $\mathbf{x}$ به شرط $T(\mathbf{x}) = t$ داده شده مستقل از θ باشد، آنگاه $\mathbf{x}$ در بردارنده اطلاعات درباره θ در مجموعه A_t نیست، یعنی اینکه $\mathbf{x}$ یک آماره بر روی مجموعه A_t است و به این معنی است که آماره $T(\mathbf{x})$ در بردارنده اطلاعات درباره θ به اندازه $\mathbf{x}$ است. در این حالت آماره $T(\mathbf{x})$ آماره بسنده است.

تعریف آماره $T(\mathbf{x})$ را آماره بسنده برای پارامتر θ گوئیم اگر تابع احتمال شرطی $P(\mathbf{X} \in A_t | T(\mathbf{X}) = t)$ ، $\forall t \in A_t$ مستقل از θ باشد.

مثال ۱-۱ فرض کنید $\mathbf{x} = (X_1, X_2, X_3, X_4)$ یک نمونه تصادفی از توزیع برنولی با پارامتر θ باشد. پس از مشخص کردن فضای نمونه نشان دهید که آماره

$$T(\mathbf{x}) = \sum_{i=1}^4 X_i \quad \text{یک آماره بسنده و آماره } S(\mathbf{x}) = \prod_{i=1}^n X_i \quad \text{بسنده نیست.}$$

حل:

$$f(x; \theta) = \theta^x (1-\theta)^{1-x}, \quad x = 0, 1, \quad 0 < \theta < 1 \quad (1-1)$$

فضای نمونه، نمونه تصادفی $\mathbf{x} = (X_1, X_2, X_3, X_4)$ برابر است با:

$$S = \{(0,0,0,0), (0,0,0,1), (0,0,1,0), (0,1,0,0), (1,0,0,0), (0,0,1,1), (0,1,0,1), (1,0,0,1), (1,0,1,0), (1,1,0,0), (0,1,1,0), (0,1,1,1), (1,1,0,1), (1,1,1,0), (1,1,1,1)\}$$

می‌دانیم که آماره $T(\mathbf{x}) = \sum_{i=1}^4 X_i$ دارای توزیع دوجمله‌ای با پارامتر ۴ و θ است،

به طوری که

$$g(t; \theta) = \binom{4}{t} \theta^t (1-\theta)^{4-t}, \quad t = 0, 1, 2, 3, 4, \quad 0 < \theta < 1$$

x		$(1, \circ, 1, 1)$	$(1, 1, \circ, 1)$	$(1, 1, 1, \circ)$
$P(\mathbf{x} = x T(\mathbf{x}) = 3)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

و سرانجام

$$P(\mathbf{x} = (1, 1, 1, 1) | T(\mathbf{x}) = 4) = \frac{P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, T(\mathbf{x}) = 4)}{P(T(\mathbf{x}) = 4)}$$

$$= \frac{f^*(1)}{\theta^*} = \frac{\theta^*}{\theta^*} = 1$$

چون تمام توابع چگالی شرطی $\mathbf{x} = x$ به شرط $T(\mathbf{x}) = t$ مستقل از θ است، پس
 آماره $T(\mathbf{x}) = \sum_{i=1}^k X_i$ آماره بسنده است. برای اینکه نشان داده شود که
 یک آماره بسنده نیست، کافی است برای افزایش نشان داد که احتمال
 $S(\mathbf{x}) = \prod_{i=1}^n X_i$ مستقل از θ نیست. برای این منظور فرض کنید $\mathbf{x} = (1, \circ, 1, 1)$ باشد.
 در این صورت $S(\mathbf{x}) = \circ$ است. به طوری که

$$P(\mathbf{x} = (1, \circ, 1, 1) | S(\mathbf{x}) = \circ) = \frac{P(X = (1, \circ, 1, 1), S(\mathbf{x}) = \circ)}{P(S(\mathbf{x}) = \circ)} = \frac{P(X = (1, \circ, 1, 1))}{P(S(\mathbf{x}) = \circ)}$$

$$= \frac{P(X_1 = 1, X_2 = \circ, X_3 = 1, X_4 = 1)}{1 - P(\mathbf{x} = 1)}$$

$$= \frac{f(1)f(\circ)f(1)f(1)}{1 - (f(1))^4} = \frac{f(\circ)f^*(1)}{1 - (f(1))^4}$$

$$= \frac{(1 - \theta)\theta^*}{1 - \theta^*} = \frac{\theta^*}{(1 + \theta)(1 + \theta^*)}$$

چون تابع احتمال شرطی تابعی از θ است، پس آماره $S(\mathbf{x})$ بسنده نیست.

مثال ۱-۲ فرض کنید متغیر تصادفی X_i دارای توزیع دوجمله‌ای با پارامترهای n_i و θ باشد. اگر آماره $T(\mathbf{x})$ به صورت

$$T(\mathbf{x}) = X_1 + X_2 + \dots + X_k$$

تعریف شده باشد. نشان دهید که $T(\mathbf{x})$ آماره بسنده برای پارامتر θ است.

حل: می‌دانیم که $X_i \sim b(n_i, \theta)$ است، بنابراین

$$T(\mathbf{x}) \sim b\left(\sum_{i=1}^k n_i, \theta\right), \quad g(t; \theta) = \binom{m}{t} \theta^t (1-\theta)^{m-t}, \quad m = \sum_{i=1}^k n_i$$

شرط اینکه آماره $T(\mathbf{x})$ بسنده برای θ باشد آن است که تابع چگالی شرطی

$P(\mathbf{x} = x | T(\mathbf{x}) = t)$ مستقل از θ باشد. اما

$$\begin{aligned} P(\mathbf{x} = x | T(\mathbf{x}) = t) &= \frac{P(\mathbf{x} = x, T(\mathbf{x}) = t)}{P(T(\mathbf{x}) = t)} \\ &= \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k, T(\mathbf{x}) = t)}{\binom{m}{t} \theta^t (1-\theta)^{m-t}} \\ &= \frac{\prod_{i=1}^k P(X_i = x_i)}{\binom{m}{t} \theta^t (1-\theta)^{m-t}} = \frac{\prod_{i=1}^k \binom{n_i}{x_i} \theta^{x_i} (1-\theta)^{n_i - x_i}}{\binom{m}{t} \theta^t (1-\theta)^{m-t}} \\ &= \frac{\theta^{\sum_{i=1}^k x_i} (1-\theta)^{\sum_{i=1}^k n_i - \sum_{i=1}^k x_i} \prod_{i=1}^k \binom{n_i}{x_i}}{\binom{m}{t} \theta^t (1-\theta)^{m-t}} \\ &= \frac{\theta^t (1-\theta)^{m-t} \prod_{i=1}^k \binom{n_i}{x_i}}{\binom{m}{t} \theta^t (1-\theta)^{m-t}} = \frac{\prod_{i=1}^k \binom{n_i}{x_i}}{\binom{m}{t}} \quad (2-1) \end{aligned}$$

چون نتیجه حاصل مستقل از θ است، آماره $T(\mathbf{x})$ برای θ بسنده است. ذکر این

نکته ضروری است که آماره بسنده منحصر به فرد نیست، چون با توجه به تعریف آماره

بسنده اگر $g(\cdot)$ تابعی یک به یک از آماره $T(\mathbf{x})$ باشد، می‌توان نشان داد که $g(T(\mathbf{x}))$ نیز آماره بسنده است. با توجه به تعریف آماره بسنده، آماره بسنده مینیمال به صورت زیر تعریف می‌شود.

تعریف آماره $T(\mathbf{x})$ را بسنده مینیمال گویند، اگر بسنده باشد و تابعی از هر آماره بسنده دیگر باشد.

۱. توزیع شرطی به دست آمده در تعریف آماره بسنده گاهی پیچیده است، علی‌الخصوص وقتی که نمونه تصادفی $\mathbf{x}$ دارای توزیع پیوسته باشد.
۲. تعریف آماره بسنده به راحتی این اجازه را به ما نمی‌دهد که آماره بسنده را مشخص کنیم. شرط لازم و کافی برای بسنده بودن آماره $T(\mathbf{x})$ ، وقتی که تابع چگالی توأم نمونه تصادفی $\mathbf{x}$ داده شده باشد، توسط نیمین تحت عنوان معیار تجزیه نیمین به صورت قضیه ۱-۱ ارائه شده است.

قضیه ۱-۱ فرض کنید نمونه تصادفی $\mathbf{x} = (X_1, X_2, \dots, X_n)$ دارای تابع چگالی توأم $f(\mathbf{x}; \theta)$ باشد آماره $T(\mathbf{x})$ آماره بسنده برای θ است اگر و فقط اگر

$$f(\mathbf{x}; \theta) = g(T(\mathbf{x}); \theta)h(\mathbf{x})$$

برای رؤیت اثبات به کتب مختلف از جمله آمار ریاضی Knight، صفحه ۱۸۳ مراجعه کنید.

مثال ۱-۳ فرض کنید $\mathbf{x} = (X_1, X_2, \dots, X_n)$ یک نمونه تصادفی از توزیع یکنواخت با پارامتر θ باشد. نشان دهید آماره مرتبه n -ام یا $X_{(n)}$ آماره بسنده برای θ است.

حل:

$$f(x; \theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta \quad \text{یا} \quad f(x; \theta) = \frac{1}{\theta} I_{[0, \theta]}(x)$$

$$\begin{aligned}
 f(x; \theta) &= \frac{1}{\theta^n}, \quad 0 \leq x_i \leq \theta \\
 &= \frac{1}{\theta^n} I(0 \leq x_1, x_2, \dots, x_n \leq \theta) = \frac{1}{\theta^n} I\left(\max_{1 \leq i \leq n} x_i \leq \theta\right) I\left(\min_{1 \leq i \leq n} x_i \geq 0\right) \\
 &= g(\max(x_1, x_2, \dots, x_n); \theta) h(\mathbf{x}) \\
 &\text{بنابراین } X_{(n)} = \max(X_1, X_2, \dots, X_n) \text{ آماره بسنده برای } \theta \text{ است.}
 \end{aligned}$$

۱-۳ کامل بودن

در این بخش با توجه به تعریف آماره بسنده و بسنده مینیمال که در بخش ۱-۲ ارائه شد، کامل بودن آماره مورد بحث قرار می‌گیرد. برای این منظور اگر $\mathbf{x} = (X_1, X_2, \dots, X_n)$ یک نمونه تصادفی از تابع چگالی احتمال $f(x; \theta)$ باشد. برای تعریف کامل بودن آماره $T(\mathbf{x}) = T(X_1, X_2, \dots, X_n)$ که تابعی از نمونه تصادفی است؛ فرض کنید $\{f(t; \theta), t \in A_t\}$ خانواده تابع چگالی آماره $T(\mathbf{x})$ باشد، آماره $T(\mathbf{x})$ را کامل گویند اگر $E_\theta(g(T(\mathbf{x}))) = 0$ آنگاه به ازای هر تابع اندازه‌پذیر g و هر θ ، $P_\theta[g(T(\mathbf{x})) = 0] = 1$.

مثال ۱-۴ فرض کنید $\mathbf{x} = (X_1, X_2, \dots, X_n)$ یک نمونه تصادفی از توزیع پواسن با

پارامتر λ باشد. نشان دهید آماره $T(\mathbf{x}) = \sum_{i=1}^n X_i$ کامل است.

حل:

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

چون $X_1, X_2, \dots, X_n$ یک نمونه تصادفی از توزیع پواسن با پارامتر λ است

نتیجه می‌شود که $T(\mathbf{x}) = \sum_{i=1}^n X_i$ دارای توزیع پواسن با پارامتر $n\lambda$ است، بنابراین

$$T(\mathbf{x}) = \sum_{i=1}^n X_i \sim P(n\lambda) \quad , \quad g(t; \lambda) = \frac{e^{-n\lambda} (n\lambda)^t}{t!} \quad , \quad t = 0, 1, 2, \dots$$

برای اینکه نشان دهیم که $T(\mathbf{x})$ آماره کامل است، برای هر $\lambda > 0$ ،

$$P[g(T(\mathbf{x})) = 0] = 1 \quad \text{نتیجه می‌گیریم که} \quad E[g(T(\mathbf{x}))] = 0$$

$$E(g(T(\mathbf{x}))) = \sum_{t=0}^{\infty} g(t; \lambda) \frac{e^{-n\lambda} (n\lambda)^t}{t!} = 0 \quad , \quad \lambda > 0$$

$$\Rightarrow \sum_{t=0}^{\infty} \frac{g(t; \lambda) (n\lambda)^t}{t!} = 0 \Rightarrow \frac{g(0)}{1} + \frac{g(1)(n\lambda)}{1} + \dots + \frac{g(k)(n\lambda)^k}{k!} + \dots = 0$$

شرط اینکه رابطه اخیر برابر با صفر باشد، لازم است $g(0), g(1), \dots, g(k), \dots$ همگی برابر با صفر شوند. چون که $(n\lambda)$ با هر توانی مخالف صفر است. بنابراین نتیجه می‌شود که برای هر $\lambda > 0$ ، $g(t) = 0$ ، در نتیجه

$$P[g(T(\mathbf{x})) = 0] = 1$$

مثال ۵-۱ فرض کنید متغیر تصادفی X دارای تابع چگالی زیر باشد.

$$P(X = 0 | \theta) = \theta \quad , \quad P(X = 1 | \theta) = 2\theta \quad , \quad \theta \in \left(0, \frac{1}{3}\right) \quad , \quad p(x = 2 | \theta) = 1 - 3\theta$$

اگر $X_1, X_2, \dots, X_n$ یک نمونه تصادفی از تابع چگالی احتمال فوق باشد و متغیر T_i به صورت زیر تعریف شده باشد.

تعداد X_i های برابر با i $T_i = i$

الف- نشان دهید که T یک آماره بسنده است.

ب- نشان دهید خانواده آماره T کامل است.

حل: الف) فرض کنید t_1, t_2 و t_3 به ترتیب تعداد X_i های برابر با صفر، یک و دو باشد. تابع درستنمایی برابر است با: