دانشکده علوم پایه و کشاورزی
مرکز تهران شرق

عنوان :

برآورد بیزی مدل وایبل نمایی شده تحت داده‌های سانسور فزاینده نوع II

پایان‌نامه برای دریافت درجه کارشناسی ارشد
در رشته ی آمار ریاضی

نگارنده:
علی بری خجسته

استاد راهنما:
دکتر پرویز نصیری

استاد مشاور:
دکترعلی شادرخ

تیر ۱۳۹۲

بسم الله الرحمن الرحیم

(گواهی اصالت، نشر و حقوق مادی و معنوی اثر)

اینجانب   دانشجوی ورودی سال   مقطع کارشناسی ارشد رشته   گواهی می‌نمایم چنان چه در پایان نامه خود از فکر، ایده و نوشته دیگری بهره گرفته‌ام با نقل قول مستقیم یا غیرمستقیم منبع و ماخذ آن را نیز در جای مناسب ذکر کرده ام .بدیهی است مسئولیت تمامی مطالبی که نقل‌قول دیگران نباشد بر عهده خویش می‌دانم و جوابگوی آن خواهم بود.

دانشجو تأیید می‌نماید که مطالب مندرج دراین پایان نامه (یا رساله) نتیجه تحقیقات خودش می باشد و در صورت استفاده از نتایج دیگران مرجع آن را ذکر نموده است.

نام و نام خانوادگی دانشجو:

تاریخ و امضاء:

اینجانب   دانشجوی ورودی سال   مقطع کارشناسی ارشد رشته   گواهی می‌نمایم چنانچه براساس مطالب پایان‌نامه خود اقدام به انتشار مقاله، کتاب، و... نمایم ضمن مطلع نمودن استاد راهنما، با نظر ایشان نسبت به نشر مقاله، کتاب، و ... و به صورت مشترک و با ذکر نام استاد راهنما مبادرت نمایم.

نام و نام خانوادگی دانشجو:

تاریخ و امضاء:

(کلیه حقوق مادی مترتب از نتایج مطالعات ، آزمایشات و نو آوری ناشی از تحقیق موضوع این پایان نامه متعلق به دانشگاه پیام نور می باشد.)

تقدیم به

پدر و مادر عزیزم به خاطر زحمات بی دریغشان

تشکر و قدردانی

سپاس ستایش خداوندی که پروردگار جهانیان است، خداوند بزرگ را شاکرم که در عرصه‌ی حیات، این حقیر را در محضر علم و دانش قرار داد و زندگی ام را به سوی علم، رهنمون ساخت تا بلکه بتوانم عظمتش را از منظر دیگری شاهد باشم.

در پایان کار بر خود لازم می‌دانم تا از همه عزیزانی که مرا در این راه یاری نموده‌اند تشکر و قدردانی نمایم.

از استاد راهنمای بسیار بزرگوار و ارجمندم جناب آقای دکتر نصیری که در تمام مراحل تدوین این پایان نامه مشعوق بنده بوده و از راهنمایی‌های ارزشمند و از لطف و مرحمت خودشان مرا برخوردار نموده‌اند نهایت تشکر و قدردانی را دارم.

از استاد گرانقدر جناب آقای دکتر شادرخ که به عنوان مشاور مرا در نیل به این مقصد همراهی نموده‌اند و با راهنماییهای بی نظیر و ارزشمند خویش مرا راهنمایی نموده‌اند کمال تشکر را دارم.

از استاد محترم و بزرگوار جناب آقای دکتر جباری که زحمت بازخوانی و اصلاح پایان نامه و قبول داوری دفاعیه را تقبل نموده‌اند کمال تشکر و قدردانی را دارم.

علی بری خجسته

تیر ماه ۹۲

چکیده

مدل وایبل نمایی شده در تحلیل قابلیت اعتماد به کار گرفته می‌شود. توزیع وایبل نمایی شده با اسامی توزیع وایبل تعمیم یافته نیز شناخته می‌شود. توزیع وایبل نمایی شده مانند توزیع نمایی، گاما، دارای دو پارامترشکل و مقیاس است.

در این پایان‌نامه پس از بررسی اولیه و ویژگی‌های این توزیع، با سانسور فزاینده نوع Π وآماره‌های آن بیشتر آشنامی‌شویم ودر ادامه به استنباط در مورد توزیع توزیع وایبل برای داده‌های سانسور شده در دو حالت مختلف (پارامتر شکل معلوم و هر دو پارامتر نامعلوم) پرداخته و در آخر برآورد بیزی مناسب برای وایبل توزیع وایبل نمایی شده بدست آورده می‌شود که برآورد بیزی به روش تقریب لیندلی محاسبه می‌شود و در پایان مقایسه‌ای با برآورد درست‌نمایی ماکزیمم این توزیع انجام داده وبرآوردگر کارا معرفی می‌شود.

کلمات کلیدی: برآورد بیزی ؛برآورد ماکزیمم درست‌نمایی ؛وایبل نمایی شده ؛سانسور فزاینده ؛تقریب لیندلی ؛ تابع زیان مربع خطا ؛تابع زیان لاینکس

فهرست مطالب

عنوان ــ **صفحه**

فهرست جداول ، نمودارها و اشکال

پیش‌گفتار

در بازار رقابتی جهان امروز که محصولات بسیار متنوع و با کیفیت‌های بالا روانه بازار می‌شوند، تولید کننده‌ها برای آن‌که بتوانند در رقابت با سایر رقبا سهم خود را از بازارهای جهانی افزایش دهند یا حداقل آن را حفظ نمایند، مجبور به اتخاذ سیاست‌های متنوع در جهت جلب هر چه بیش‌تر رضایت مشتریان خود هستند. علاوه بر فنون بهبود و کنترل کیفیت یکی سیاست‌هایی که کارخانه تولید کننده محصولات برای جلب رضایت مشتریان خود به کار می‌بندد، فراهم کردن ضمانت برای طول عمر محصولات (گارانتی) است. به طور قطع ضمانت طول عمر محصولات هزینه‌های اضافی را برای تولید کنندگان به وجود خواهد آورد.

بخش عمده‌ای از این هزینه‌ها مربوط به تعمیر و یا تعویض قطعاتی است که قبل از اتمام دور هضم انتخاب می‌شوند و هم چنین یک نکته بسیار مهم این است که یک محصول از دید تولید کننده لازم نیست خیلی بیش از میزان طول عمر ضمانت شده، عمرکند! زیرا می‌بایست نیاز به معرف در جامعه باقی بماند. بنابراین تعیین هر چه دقیق‌تر طول عمر محصولات بسیار ضروری می‌نماید و لازم است محصولات قبل از این‌که وارد بازار شوند تحت آزمایش‌های مربوط به طول عمر قرار بگیرند تا تولید کننده بتواند هزینه‌ها را برآورد کند و توزیع زمان شکست محصولات خود را بداند. البته اطلاعاتی که درطول این آزمایش‌ها به دست می‌آیند در بسیاری از موارد دیگر می‌توانند مفید واقع شوند. آزمایش‌های مربوط به طول عمرمحصولات می‌توانند در جهت بهبود کیفیت و هم چنین افزایش توان رقابتی محصولات کمک کند.

در بسیاری از مطالعات مربوط به طول عمر با مواردی مواجه می‌شویم که واحدهای آزمایشی قبل از اتمام آزمایش از مطالعه خارج می‌شوند. این حذف شدن ممکن است به صورت غیر عمدی و یا از قبل توسط آزمایش‌گر طراحی شده باشد. این حذف شدن ممکن است به صورت غیر عمدی ویا از قبل توسط آزمایش‌گر طراحی شده باشد. به عنوان مثال، اگر یکی از واحدهای آزمایش به طور تصادفی خراب شود. (البته به دلیلی غیر از پایان عمر) یا شخصی تحت مطالعه از ادامه همکاری کناره‌گیری کند، حذف کند، حذف از نوع غیر عمری رخ داده، ولی اگر به علت برخی شرایط ویژه مانند تمام شدن بودجه و امکانات یا به پایان رسیدن زمان در نظر گرفته شده برای طرح، آزمایش‌گر، آزمایش را قبل از حصول زمان خرابی تمام واحدها متوقف کند حذف از نوع عمری اتفاق افتاده است و آزمایشگر آزمایش را براساس یک طرح سانسور مشخص تعریف می‌نماید. به آن داده‌هایی که قبل از مشاهده زمان شکست آن‌ها از آزمایش حذف می‌شوند، داده‌های سانسور شده می‌گویند.

در این پایان‌نامه در فصل اول کلیاتی در مورد مطالب ارائه می‌شود در فصل دوم خواص و گشتاورهای آماره‌های ترتیبی سانسور فزاینده بحث می‌شود در ادامه در فصل سوم برآورد درستنمایی و بیزی برای توزیع وایبل و در فصل چهارم برآورد بیزی و برآورد درست نمایی برای توزیع وایبل نمایی شده تحت داده‌های سانسور فزاینده نوع دو ارائه می‌شود ودر پایان با مثال عددی شبیه‌سازی انجام‌می‌شود.

در سال‌های اخیر که آزمون‌های طول عمر در صنعت گسترش بسیار زیادی پیدا کرد و روش‌های نمونه‌گیری مختلفی ارائه شد از آن‌جا که می‌بایست برای هر روشی نمونه‌گیری یک روش استنباط مناسب ارائه شود. مقالات استنباط آماری زیادی در این زمینه به چاپ رسیده است. از این میان می‌توان به افرادی که فعالیت بیش‌تر و موثرتری داشته‌اند اشاره کرد.

در مورد سانسور معمولی نوع I و نوع II [1] و استنباط‌های مربوط به آن‌ها مقالات متعددی در ادبیات این موضوع وجود دارد که برای مثال می‌توان به کوهن[2] (۱۹۹۱) و بالاکریشنان و کوهن (۱۹۹۱) اشاره کرد. همان‌طوری که قبلاً اشاره شد، در سانسورهای معمولی این اجازه وجود ندارد که واحدها در جایی قبل از نقطه پایان از آزمایش خارج شود، کوهن (۱۹۶۳ و ۱۹۶۶) یکی از اولین کسانی بود که روی طرح عمومی‌تری که طرح سانسور فزاینده[3] نامیده شده مطالعه کرد.

من[4] (۱۹۷۱) وویورس[5] و بالاکریشنان (۱۹۹۴) استنباط خطی تحت سانسور فزاینده نوع II را برای توزیع‌های وایبل و نمایی مطالعه کردند. بالاکریشنان و کانان[6] (۲۰۰۰)، بالاکریشنان و همکاران (۲۰۰۴ و ۲۰۰۳)، بالاکریشنان و اصغرزاده (۲۰۰۵) و اصغرزاده (۲۰۰۶) به ترتیب برای توزیع‌های لوجستیک، نرمال، مقادیر غایی، لجستیک و لجستیک تعمیم یافته، مسئله استنباط را برای داده‌های سانسور فزاینده نوع II مورد مطالعه قرار دادند. برای مطالعه جزئیات بیش‌تر در مورد سانسور فزاینده می‌توان به کتاب بالاکریشنان و آگاوالا (۲۰۰۰) مراجعه کرد.

[1] Progressive censoring
[2] cohen
[3] Conventional Type-I and Type- II censoring
[4] Mann
[5] Viveros and kannan
[6] Balakrishnan

فصل اول

کلیات

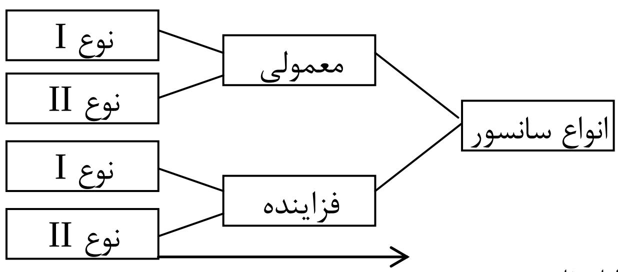
۱-۱-مقدمه

در آزمایشات طول عمر و مطالعات قابلیت اعتماد، واحدهای آزمایشی ممکن است قبل از مشاهده زمان شکست آنها، از روند آزمایش کنار گذاشته شوند یا از بین بروند. برای مثال افراد در یک آزمایش کلینیکی ممکن است از روند مطالعه خارج شوند یا مطالعه به علت فقدان بودجه زودتر خاتمه پیدا کند. در یک آزمایش صنعتی، واحدها ممکن است قبل از مشاهده زمان شکست آنها خراب شوند. در بسیاری از حالات، حذف واحدها از آزمایش، کاری از قبل برنامه‌ریزی شده است که به منظوره صرفه‌جویی در وقت و هزینه انجام می‌شود. داده‌های حاصل از این آزمایش را داده‌های سانسور شده گویند.

در این پایان‌نامه، طرح سانسور فزاینده نوع Π مورد مطالعه قرار می‌گیرد. آزمایش طول عمر n قطعه را در نظر بگیرید. در این طرح با مشاهده اولین شکست، R_1 واحد از واحدهای آزمایشی سالم باقیمانده به تصادف از آزمایشی خارج می‌شوند و در زمان دومین شکست، R_2 واحد از واحدهای آزمایشی سالم باقیمانده به تصادف از ازمایش خارج می‌شوند و این کار ادامه می‌یابد تا در زمان mامین شکست همه واحدهای باقیمانده یعنی $R_m = n - R_1 - R_2 \ldots - R_{m-1} - m$ از آزمایش خارج شوند. در این حالت زمان‌های مربوط به حذف واحدها متغیرها تصادفی خواهند بود وتعداد R_iها از قبل تعیین شده است. اگر $R_1 = R_2 = \ldots = R_{m-1} = 0$ باشد آنگاه طرح مذکور به یک طرح فاقد سانسور (طرح نمونه کامل) تبدیل می‌شود. اگر $R_1 = R_2 = \ldots = R_{m-1} = 0$ و $R_m = n - m$ آن‌گاه این طرح به سانسور معمولی نوع Π تبدیل می‌شود که در ان فقط اولین m آماره مرتب معمولی مشاهده شده‌اند.

۱-۲-انواع مختلف سانسور

در شکل (۱) نمودار تفکیک انواع سانسورهای پرکاربرد آورده شده است. در ادامه در مورد همه آن‌ها توضیح داده خواهد شد.



سانسور مورد بررسی در این پایان نامه

شکل (۱) نمودار تفکیک انواع سانسور

۱-۲-۱- سانسور معمولی نوع I

یک آزمایش طول عمر را در نظر بگیرید که در آن n واحد، در زمان صفر در معرض آزمایش طول عمر قرار می‌گیرد. اگر آزمایش فقط تا زمان مشخص و از پیش تعیین شده $\mathbf{t_c}$ ادامه یابد و واحدهای آزمایشی سالم از زمان $\mathbf{t_c}$ به بعد کنار گذاشته شود، در این صورت سانسور معمولی نوع I رخ می‌دهد.

در این نوع آزمایش $\mathbf{t_c}$، زمان سانسور و داده‌هایی با مقدار بیش‌تر از $\mathbf{t_c}$، داده‌هایی سانسور شده نوع I نامیده می‌شوند. در این حالت زمان اتمام آزمایش، $(\mathbf{t_c})$ ثابت ولی تعداد واحدهایی که قبل از زمان $\mathbf{t_c}$ خراب می‌شوند یک متغیر تصادفی خواهد بود. یعنی هر تعدادی از صفر تا n می‌تواند باشد و ضعف این روش در این است که ممکن است مشاهدات تا قبل از زمان $\mathbf{t_c}$، صفر یا تعداد بسیار اندکی باشد.

۱-۲-۲-سانسور معمولی نوع I

در آزمایش طول عمر n قطعه، اگر آزمایش‌گر تصمیم بگیرد آزمایش را زمانی خاتمه دهد که m واحد از واحدهای آزمایشی از کار بیافتند. در این صورت سانسور معمولی نوع II رخ می‌دهد. در این سانسور شکست‌های مشاهده شده ثابت، ولی زمان اتمام آزمایش یک متغیر تصادفی است.
مشکل اصلی این سانسور این است که ممکن است زمان بسیار زیادی طول بکشد تا mامین شکست رخ دهد که ممکن است از حوصله آزمایش خارج شود.

۱-۲-۳-سانسور فزاینده

اگر آزمایش‌گر تصمیم بگیرد واحدهای آزمایشی را در زمان‌های متفاوت یک آزماش طول عمر حذف کند یا در طول آزمایش تعدادی از واحدهای آزمایشی سالم به طور غیر عمدی در زمانی به غیر از زمان پایان آزمایش (که از قبل توسط وی مشخص شده است)، از آزمایش خارج شدند، در این صورت هیچ یک از طرح‌های سانسور که تاکنون بیان شده است، نمی‌توانند مورد استفاده قرار بگیرند. زیرا در هیچ کدام از طرح‌های مطرح شده، به آزمایش‌گر اجازه داده نمی‌شود که واحدهای سالم آزمایشی را قبل از زمان یا تعداد تعیین شده از آزمایش حذف کند.
این اجازه که آزمایش‌گر بتواند واحدهای آزمایشی را در زمان‌های مختلف آزمایش حذف کند، در مطالعات فرسایشی بسیار مطلوب است، در این مطالعات برای مشاهده زمان شکست مورد نظر لازم است واحد آزمایشی در مراحل مختلف و به صورت کامل از بین برود. هم‌چنین مواردی وجود دارد که در آن‌ها واحدهای آزمایشی سالم که از یک آزمایش حذف می‌شوند، می‌توانند برای آزمایش‌های

دیگر مورد استفاده قرار بگیرد. در این حالت نیز حذف واحدهای سالم آزمایشی در زمان‌های مختلف می‌تواند باعث کاهش چشم‌گیر در هزینه و وقت آزمایشی شود.

در ذیل مواردی که در آن‌ها استفاده از سانسور فزاینده می‌تواند مناسب باشد بیان شده است:

۱- در مواردی که برای آزمایش‌های دیگر نیاز به واحدهای آزمایشی و یا وسایل آزمایشی وجود داشته باشد.

۲- در مواردی که شکست یک واحد آزمایشی به طور ضمنی نشان دهنده طول عمر تعداد دیگری از واحدهای آزمایشی باشد.

۳- در مواردی که با توجه به ماهیت آزمایش‌های انجام شده استفاده از سایر طرح‌های سانسور مقدور نباشد.

۴- در مواردی که محدودیت بودجه و امکانات وجود داشته باشد.

در طرح‌های سانسور (فزاینده)، همانند سانسور معمولی دو نوع طرح کلی وجود دارد.

۱-۲-۳-۱-سانسور فزاینده نوع I [1]

اگر آزمایش‌گر تصمیم بگیرد در زمان‌های از قبل تعیین شده $T_۱, T_۲, ..., T_m$ به ترتیب $R_۱, R_۲, ..., R_m$ واحد را به صورت تصادفی از آزمایش حذف کند در این صورت سانسور فزاینده از نوع I رخ داده است واضح است که اگر در زمان T_i تعداد واحدهای زنده مانده کم‌تر از R_iها باشد آن‌گاه کل تعداد واحدهای باقیمانده از آزمایش حذف می‌شود. بنابراین تعداد واحدهای حذف شده در زمان T_i که آن با R_i^{obs} نشان داده می‌شود، عبارتنداز:

$$R_i^{obs} = \min(R_i , T_i \text{ تعداد واحدهای باقیمانده در زمان })$$

بنابراین آزمایش در زمان T_m با حذف R_m^{obs} واحد باقیمانده خاتمه می‌یابد. این نوع سانسور که یک حالت تعمیم یافته از سانسور نوع I معمولی است. سانسور فزاینده نوع I نامیده می‌شود.

که در آن اگر:

$$R_i = ۰ \qquad i = ۱, ۲, ..., m-۱$$

آنگاه طرح سانسور نوع I معمولی به دست می‌آید. در این طرح سانسور همانند طرح سانسور معمولی $T_۱, T_۲, ..., T_m$ ثابت و از قبل تعیین شده‌اند. ولی تعداد واحدهای آزمایشی که از بین می‌روند یک متغیر تصادفی است.

---

[1] Progressire Type- I Censoring

۱-۲-۳-۲-سانسور فزاینده نوع II [۱]

این نوع سانسور، سانسور مورد مطالعه در این پایان‌نامه است. اگر آزمایش‌گر تصمیم بگیرد با مشاهده اولین شکست R_1 واحد از واحدهای آزمایشی سالم باقیمانده را به تصادف از آزمایش خارج کند و در زمان دومین شکست R_2 واحد از واحدهای آزمایشی سالم باقیمانده را به تصادف از آزمایشی خارج کند و این کار را ادامه دهد تا در زمان mامین شکست همه واحدهای باقیمانده یعنی:

$$R_m = n - R_1 - R_2 - \ldots - R_{m-1} - m$$

را از آزمایش خارج کند، در این صورت طرح سانسور فزاینده نوع II رخ می‌دهد. شکل (۲).

در این حالت زمان‌های مربوط به حذف واحدها متغیرهای تصادفی خواهند بود. در این نوع سانسور با مشاهده i امین شکست R_i تا از واحدهای باقیمانده به صورت تصادفی از آزمایش حذف می‌شوند. همانند سانسور نوع I تعداد R_iها از قبل تعیین شده است. در نتیجه اجرای این طرح سانسور، m مقدار مرتب شده به دست می‌آید که آماره‌های ترتیبی سانسور فزاینده نوع II نامیده می‌شوند اگر:

$$R_m = n - m, \quad R_1 = R_2 = \ldots = R_{m-1} = 0$$

آنگاه این طرح به سانسور معمولی نوع II تبدیل می‌شود که در آن‌جا فقط اولین m آماره مرتب معمولی مشاهده شده‌اند. همچنین اگر $R_1 = R_2 = \ldots = R_m = 0$ باشد آن‌گاه طرح مذکور به یک طرح فاقد سانسور (طرح نمونه کامل) تبدیل می‌شود. که در آن همه n آماره مرتب معمول مشاهده می‌شوند.

فرض کنید n واحد مستقل در یک آزمایش طول عمر قرار گرفته‌اند با زمان‌های شکست متناظر $X_1, \ldots, X_n$ که دارای توزیع یکسان با تابع توزیع تجمعی $F(x)$ و تابع چگالی احتمال $f(x)$ می‌باشند. هم‌چنین فرض کند تعداد شکست‌هایی که از قبل تعیین شده‌اند، m باشد و طرح سانسور فزاینده نیز $(R_1, R_2, \ldots, R_m)$ باشد. در سانسور فزاینده نوع II زمان‌های مربوط به شکست m واحد آزمایشی با $X_{i:m:n}^{(R_1, R_2, \ldots, R_m)}$ $i = 1, 2, \ldots, m$ نشان داده می‌شود.

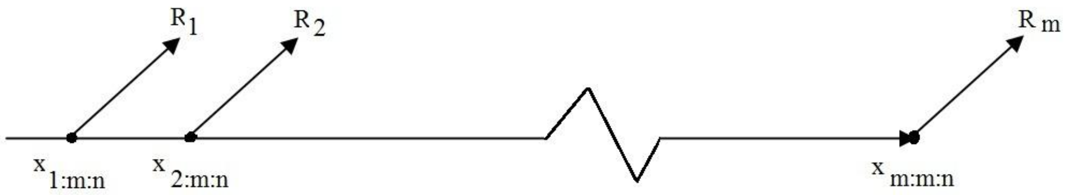
برای سادگی هرگاه طرح سانسور از قبل مشخص شده باشد، زمان‌های شکست مربوط به m واحد مذکور با $X_{i:m:n}$ $i = 1, 2, \ldots, m$ نشان داده می‌شود.

باید توجه داشت که $X_{i:n}$ ، $X_{i:m:n}$ (که دومی، i امین آماره مرتب از نمونه اصلی به حجم n است)، متفاوتند زیرا ممکن است که i امین‌مان شکست مرتب شده از نمونه اصلی قبل از مشاهده i امین

---

[۱] Progressive Type - II

آماره مرتب سانسور نوع II فزاینده حذف شده باشد. بنابراین می‌توان گفت که $\mathbf{x_{i:m:n} \geq x_{i:n}}$ و همچنین: $\mathbf{x_{1:m:n} = x_{1:n}}$



شکل (۲) : نحوه سانسور شدن واحدها در سانسور فزاینده

۱-۳-تابع زیان

پارامتر مجهول θ از یک توزیع معین را در نظر بگیرید. فرض کنید $\hat{\theta}$ برآوردگری از θ را نشان دهد. اگر D نمایانگر کلاس تمام برآوردگرها باشد و Θ فضای پارامتر مورد نظر باشد. تابعی تحت عنوان تابع زیان که آن را با L نشان می‌دهیم معرفی می‌گردد که تابعی دو متغیره از $\Theta \times D$ به زیر مجموعه‌ای از اعداد حقیقی نامنفی است. یعنی:

$$L : \Theta \times D \rightarrow R^T$$

و در شرایط زیر صدق می‌کند:

۱ـ برای همه برآوردهای ممکن $\hat{\theta}$ و هر $\theta \in \theta$، $L(\hat{\theta},\theta) \geq 0$

۲ـ برای $\hat{\theta} = \theta$، $L(\hat{\theta},\theta) = 0$.

وقتی θ مقدار واقعی پارامتر باشد، چنانچه θ با $\hat{\theta}$ برآورد شود، $L(\hat{\theta},\theta)$ برابر زیانی است که متحمل می‌شویم.

در آمار توابع زیان متعددی با توجه به نوع مسئله به کار گرفته می‌شوند. توابع زیان معروف عبارتنداز:

ـ تابع زیان مربع خطا: $L(\hat{\theta},\theta) = (\hat{\theta} - \theta)^2$

ـ تابع زیان قدرمطلق خطا: $L(\hat{\theta},\theta) = \left|\hat{\theta} - \theta\right|$

ـ تابع زیان وزنی مربع خطا: $w(\theta) \geq 0$ و $L(\hat{\theta},\theta) = W(\theta)(\hat{\theta} - \theta)^2$

ـ تابع زیان صفر و یک که بیشتر در آزمون‌های فرض به کار می‌رود:

$$L(\hat{\theta},\theta) = \begin{cases} 1 & \theta \in \Theta_i \\ 0 & \theta \in \Theta_j \end{cases} \quad i \neq j$$

و نیز تابع زیان لاینکس[۱] که به تفصیل به آن خواهیم پرداخت.

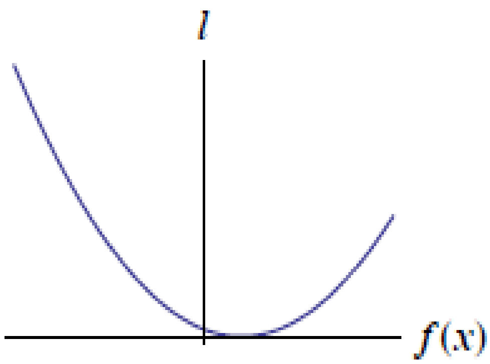
حال فرض می‌کنیم که یک تابع زیان مناسب برای مسأله برآورد تعریف کرده‌ایم و تابع زیان را به عنوان یک اندازه خطا یا زیان در نظر گرفته‌ایم. هدف انتخاب برآوردگری است که این خطا یا زیان را کوچک سازد. تابع زیان چنانکه مشهود است به برآوردگر $\hat{\theta}$ بستگی دارد و $\hat{\theta}$ تابعی از نمونه تصادفی می‌باشد. بنابراین تابع زیان به نمونه $x_1,\dots,x_n$ بستگی دارد.

نمی‌توان امیدوار بود که زیان برای هر نمونه ممکن کوچک باشد. اما می‌توان سعی کرد که به طور متوسط زیان کاهش داده شود. بنابراین اگر هدف از انتخاب برآوردگری که زیان را کوچک می‌سازد به انتخاب برآوردگری که متوسط زیان را کوچک می‌کند تغییر داده شود، می‌توان وابستگی زیان به نمونه $x_1,\dots,x_r$ حذف شود. این مفهوم در تعریف زیر گنجانده شده است.

۱-۳-۱-تابع زیان مربع خطا

تابع زیان مربع خطا؛ تابعی مناسب برای اهداف مسائل رگرسیونی است. با این وجود، دارای نقطه بحرانی یا شکاف است.

نقاط دور افتاده در داده‌ها (نقاط مجزا شده‌ای که از تابع هدف مورد نظر دور هستند) با مربع کردن خطا خیلی محسوس‌تر می‌شوند. بنابراین، داده‌ها باید ابتدا روی نقاط دور افتاده فیلتر شوند در غیر این صورت این تابع خطا مناسب نیست.



شکل (۳): نمودار تابع زیان مربع خطا

۱-۳-۲-تابع زیان لاینکس

در برآورد متوسط عمر استفاده از تابع زیان متقارن مربع خطا ممکن است مفید نباشد. تابع زیان مربع خطا وزن‌های یکسانی را به خطاهای مثبت و منفی برآورد می‌دهد اما در حقیقت بیش برآورد معمولاً جدی‌تر و خطیرتر از کم برآورد است.

---

[۱] LINEX (linear Exponertial)

در سال (۱۹۸۶) واریان[۱] و زلنر[۲] یک تابع زیان نامتقارن تحت عنوان تابع زیان لاینکس ارائه دادند. شکل تابع زیان لاینکس برای هر پارامتر θ به صورت زیر تعریف شده است.

$$L(\Delta)=e^{c\Delta}-c\Delta-1 \quad \text{و} \quad c\neq 0 \quad \text{و} \quad \Delta=\frac{\hat{\theta}}{\theta}-1 \qquad (۴\_۱)$$

که مقدار c بیانگر میزان و درجه عدم تقارن این تابع و علامت c نشانگر جهت این عدم تقارن است. در حقیقت شکل تابع زیان لاینکس که تابعی محدب است توسط مقدار و علامت c تعیین می‌شود. (شکل ۱ـ۱).

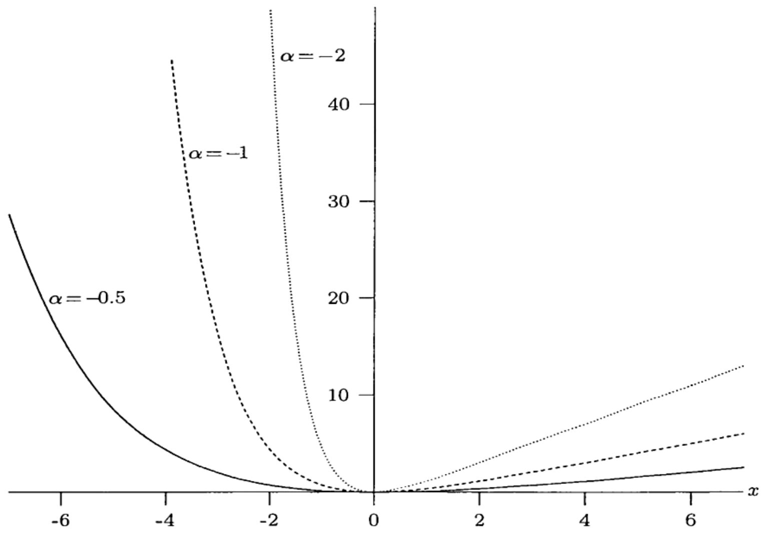
مقدار مثبت c نشان دهندهٔ این است که بیش برآورد جدی‌تر از کم برآورد است و بالعکس. بر این اساس می‌توان تصویری از تابع زیان لاینکس در نظر آورد که مثلاً برای $c<0$ در صورتی که خطای برآورد $0<(\hat{\theta}-\theta)$ باشد، تابع تقریباً به صورت نمایی صعود می‌کند و در صورتی که خطای برآورد $0>(\hat{\theta}-\theta)$ باشد، شکل تابع تقریباً به صورت خطی می‌باشد. برای c=۱ تابع زیان کاملاً نامتقارن است و برای مقادیر کوچک $|c|$ (مقادیر نزدیک به صفر) تابع زیان لاینکس تقریباً متقارن و نزدیک به تابع زیان مربع خطا است. چنان که مشاهده می‌شود با جایگذاری بسط تیلور[۳] تابع $e^{c\Delta}$ در تابع $L(\Delta)$ به نتیجه گفته شده خواهیم رسید.

$$\begin{cases} L(\Delta)=e^{c\Delta}-c\Delta-1 \\ e^{c\Delta}\approx 1+c\Delta+\dfrac{c^2\Delta^2}{2} \end{cases} \Rightarrow \; L(\Delta)\approx 1+c\Delta+\frac{c^2\Delta^2}{2}-c\Delta-1$$

و چون c بسیار کوچک است نتیجه گرفته می‌شود که تابع زیان متقارن مربع خطا حالت خاصی از تابع زیان نامتقارن لاینکس است.

---

۱ Varian
۲ Zellner
۳ Taylor



$Exp\{\alpha x\} - \alpha x - 1$

شکل (۴): نمودار تابع زیان لاینکس

۱-۴-تابع مخاطره[۱]

برای تابع زیان مفروض $L(\hat{\theta},\theta)$ تابع مخاطره برآوردگر $\hat{\theta}$ با $Risk(\hat{\theta})$ نشان داده می‌شود و به صورت زیر تعریف می‌گردد:

$$Risk(\hat{\theta}) = E(L(\hat{\theta},\theta))$$

حال چنانچه از توابع زیان معرفی شده در بالا امید ریاضی گرفته شود، توابع مخاطره مربوطه به دست می‌آیند. به طور ویژه اگر از تابع زیان مربع خطا یعنی:

$$\mathrm{L}(\hat{\theta},\theta) = (\hat{\theta}-\theta)^2 \qquad (۱\_۲)$$

امید ریاضی گرفته شود خواهیم داشت:

$$\mathrm{Risk}(\hat{\theta}) = \mathrm{E}\left((\hat{\theta}-\theta)^2\right) \qquad (۱\_۳)$$

که همان میانگین مربع خطای آشناست و با MSE نشان داده می‌شود.

توجه شود که به طور کلی برآوردگر به طور یکنواخت با کوچکترین مخاطره وجود ندارد. روش، در یافتن برآوردگر مطلوب بر این اساس است که برآوردگرها با در نظر گرفتن ماکزیمم مخاطره‌های مربوطه با هم مقایسه شوند که قاعدتاً برآوردگری مطلوب‌تر است که دارای کوچکترین ماکزیمم مخاطره باشد.

Risk function[۲]