دانشکده علوم ریاضی

پایان‌نامه جهت اخذ درجه کارشناسی ارشد

آمار ریاضی

برآورد پارامترهای توزیع نمایی تعمیم یافته تحت داده‌های سانسور شده

استاد راهنما

دکتر ملیحه عباس نژاد

استاد مشاور

دکتر آرزو حبیبی راد

نگارش

علی عباسی فرد

شهریور ۱۳۹۰

تقدیم به:

اسطوره های زندگیم، پناه خستگیم و امید بودنم

پدر و مادر عزیزم

و

تقدیم به برادرم:

که همواره در طول تحصیل متحمل زحماتم بود و تکیه گاه من در مواجهه با مشکلات، و وجودش مایه دلگرمی من می باشد.

پیش‌گفتار

توزیع نمایی تعمیم یافته از جالب ترین تعمیم های توزیع نمایی است که توسط گوپتا و کاندو (۱۹۹۹) معرفی شد و یک توزیع مهم در تحلیل بقا به حساب می آید. این توزیع در برخی حالات بهتر از توزیع گاما و وایبل به داده ها برازش داده می شود و از انعطاف پذیری بیشتری برخوردار است. همچنین توزیع نمایی تعمیم یافته معایب مهم توزیع های گاما و وایبل را ندارد، برای مثال از معایب مهم توزیع گاما این است که تابع توزیع و تابع بقای آن، زمانی که پارامتر شکل آن یک عدد صحیح نباشد به آسانی قابل محاسبه نیست و از معایب توزیع وایبل می توان به نداشتن بسیاری از ویژگی های ترتیب های تصادفی، اشاره کرد. از این رو اخیراً این توزیع مورد توجه بسیاري از محققان قرار گرفته است.

از آنجا که سانسور یک پدیده متداول در آزمون های طول عمر و مطالعات قابلیت اعتماد می باشد بنابراین تحلیل داده های سانسور شده به علت وسعت کاربرد و اهمیت بررسی این نوع داده ها، در بسیاری از تألیفات آماری به چشم می خورد. یکی از مهم ترین روش های استنباط آماری در مدل سازی داده های سانسور شده بحث برآورد پارامتر است. تاکنون مسأله برآورد براساس داده های سانسور شده از جنبه های مختلف مورد بررسی قرار گرفته است و روش های متعددی نیز برای این منظور پیشنهاد شده است.

با توجه به اهمیت و کاربرد توزیع نمایی تعمیم یافته در تحلیل بقا، در سال های اخیر استنباط بر اساس داده های سانسور شده در این توزیع مورد توجه بیشتری قرار گرفته و برمبنای این داده ها برآوردگرهای مختلفی برای این توزیع ارائه شده است.

مطالب این پایان نامه با توجه به موضوع و محتوی در فصول زیر گرد آوری شده است:

- در فصل اول، ابتدا به معرفی مختصری از قابلیت اعتماد و کمیت های مهم آن می پردازیم و سپس کلیاتی درباره داده های سانسور شده و انواع آن بیان می کنیم و همچنین مفاهیم و مقدمات مورد نیاز فصل های بعد که شامل ترتیب های تصادفی ، خواص مجانبی برآوردگرهای درستنمایی ماکسیمم ، روش تکرار عددی و الگوریتم EM است، را ارائه می دهیم.

- فصل دوم اختصاص به معرفی خانواده نمایی تعمیم یافته و بررسی برخی از ویژگی های آن دارد. در این فصل رفتار تابع نرخ خطر توزیع نمایی تعمیم یافته با توزیع های گاما و وایبل مقایسه شده و گشتاورها، آنتروپی و توزیع مقادیر فرین توزیع نمایی تعمیم یافته مورد مطالعه قرار گرفته است. در ادامه به بررسی ترتیب‌های تصادفی در این توزیع پرداخته ایم و همچنین برآوردگرهای درستنمایی ماکسیمم پارامتر ها و فواصل اطمینان مجانبی آن ها را براساس داده های کامل به دست می آوریم.

- در فصل سوم، برآوردگرهای درستنمایی ماکسیمم و فواصل اطمینان مجانبی را تحت داده های سانسور شده از چپ به دست می آوریم. در انتها شبیه سازی گسترده ای جهت مشاهده رفتار روش های پیشنهاد شده، صورت گرفته است.

- در فصل چهارم به بررسی استنباط آماری توزیع نمایی تعمیم یافته تحت داده های سانسور فزاینده می پردازیم. در ابتدا جزئیاتی از مطالب فصل را آورده ایم و سپس با استفاده از الگوریتم EM برآورد گرهای درستنمایی ماکسیمم را محاسبه کرده و همچنین ماتریس اطلاع فیشر مشاهدات را با استفاده از اصل اطلاع گم شده به دست می آوریم و در ادامه از آن برای محاسبه فواصل اطمینان مجانبی و احتمالات پوشش استفاده می کنیم. به کمک نتایج شبیه سازی رفتار برآوردگرهای درستنمایی ماکسیمم در طرح های نمونه گیری مختلف بررسی می‌کنیم و همچنین برای درک بهتر مطالب، مثالی با داده های واقعی ارائه می‌دهیم.

باید اشاره کرد که مقاله ای تحت عنوان " بررسی ترتیب های تصادفی در خانواده نمایی تعمیم یافته " از این پایان نامه استخراج و در هشتمین سمینار احتمال و فرایندهای تصادفی که در دانشگاه گیلان برگزار

گردید و ارائه شد. همچنین مقاله ای مستخرج از پایان نامه تحت عنوان:" استنباط توزیع نمایی تعمیم یافته تحت داده های سانسور از چپ" در دومین کارگاه قابلیت اعتماد که در دانشگاه اصفهان برگزار شد، ارائه گردید.

در پایان لازم می دانم که تشکر و قدردانی خالصانه خود را نسبت به استاد عزیز و ارجمندم ، خانم دکتر عباس نژاد که زحمت راهنمایی این پایان نامه را برعهده داشتند و همچنین خانم دکتر حبیبی راد که سمت مشاوره این پایان نامه را عهده دار بودند، اعلام نمایم. از داوران محترم جناب آقایان دکتر رزمخواه و دکتر محتشمی تشکر می نمایم. همچنین از همکاری های صمیمانه مسئولین کتابخانه دانشکده علوم ریاضی که در تهیه مراجع و کتاب های مورد نیاز اینجانب را یاری رساندند، کمال تشکر را دارم. در پایان، از زحمات، صبر و شکیبایی بی دریغ پدر و مادر و همسر مهربانم، سپاسگزارم.

علی عباسی فرد

شهریور ۱۳۹۰

فهرست مطالب

لیست تصاویر

لیست جداول

فصل ۱

مفاهیم و مقدمات

۱.۱ مقدمه ای بر قابلیت اعتماد

در دنیای مدرن امروز قابلیت اعتماد نقش چشم گیری در بالا بردن استاندارد و کیفیت تجهیزات و تولیدات صنعتی ایفاء می‌کند. قابلیت اعتماد کلمه ای با تعابیر متفاوت است. هنگامی که در ارتباط با یک انسان به کار می رود به معنی توانایی فرد در انجام وظیفه ای است که به وی محول می شود. این تعبیر می تواند به قطعه الکتریکی، یک محصول صنعتی و یا یک سیستم متشکل از اجزاء تعمیم داده شود؛ بدین معنی که توانایی انجام آنچه که به سیستم محول شده است که باید انجام دهد را قابلیت آن سیستم می گویند.
از نقطه نظر ریاضی و آمار قابلیت اعتماد یک احتمال است. فرض کنید که یک سیستم مورد نظر باشد. آنگاه قابلیت اعتماد سیستم به این صورت تعریف می شود: احتمال اینکه سیستم در یک بازه زمانی معین و تحت شرایط مشخص بتواند بدون شکست کار کند.
فرض کنید سیستم دارای طول عمر T باشد. طبیعی است که T یک متغیر تصادفی است که می تواند در فاصله $(۰,\infty)$ مقدار بگیرد. T می تواند پیوسته یا گسسته باشد. به عبارت دیگر سیستم می تواند طول عمری داشته باشد که هر مقدار را در $(۰,\infty)$ اختیار کند یا گسسته باشد که مقادیر شمارش پذیر را در فاصله $(۰,\infty)$ اختیار کند. اگر T دارای تابع چگالی احتمال f باشد آنگاه از نقطه نظر ریاضی و آمار قابلیت اعتماد سیستم در زمان t که آن را با $S(t)$ نمایش می دهیم، عبارتست از

$$S(t) = P(T > t) = \int_t^\infty f(x)dx. \tag{۱.۱}$$

تخمین $S(t)$ که آن را تابع بقا یا تابع قابلیت اطمینان نیز می‌نامند، یکی از اهداف مهم مطالعات قابلیت اعتماد است. به موازات مطالعه روی $S(t)$ معمولاً یکی دیگر از اندازه های مهم در قابلیت اعتماد که مبنای تحلیل بسیاری از داده های قابلیت اعتماد است اندازه ای به نام نرخ خطر می باشد که با استفاده از $S(t)$ تعریف می‌شود. اگر نرخ خطر در زمان t را با $h(t)$ نمایش دهیم آنگاه به صورت زیر تعریف می شود

$$\begin{aligned} h(t) &= \lim_{\delta \to \infty} P(t < T < t + \delta | T > t) \\ &= \frac{f(t)}{S(t)}. \end{aligned} \tag{۲.۱}$$

$h(t)$ نرخ از کار افتادگی یک سیستم با طول عمر T را در زمان t نشان می دهد، با فرض این که در زمان t سیستم هنوز کار می‌کرده است.

رفتار $h(t)$ در تحلیل داده های طول عمر مهم است. برای مثال اگر $h(t)$ رفتار صعودی داشته باشد آنگاه سیستم براثر گذشت زمان دچار فرسودگی و سالخوردگی شده و قابلیت اعتماد شرطی برای آن با گذشت زمان کاهش پیدا می‌کند. همچنین برای دو سیستم با نرخ خطر های h_1 و h_2، اگر نرخ خطر h_1 همواره کمتر از نرخ خطر h_2 باشد، آنگاه می‌توان نتیجه گرفت که سیستم اول قابل اعتمادتر از دومی است.
کمیت مهم و اصلی دیگر در تحلیل داده های قابلیت اعتماد، تابع میانگین باقیمانده عمر در زمان t است. برای هر فرد با طول عمر t، این تابع میزان انتظار از باقیمانده طول عمر او را اندازه گیری می‌کند. تابع

$$mrl(t) = E(T - t|T > t).$$

را تابع میانگین طول عمر باقیمانده در زمان t می نامند.
از خصوصیات متمایز داده‌های قابلیت اعتماد نیز می‌توان به این مورد اشاره کرد که داده های قابلیت اعتماد مثبت هستند. این به این معنی است که با متغیرهای تصادفی منفی و یا توزیع هایی که تکیه گاه آن ها مقادیر منفی را اختیار کند، سر و کار نداریم.
از متداول ترین توزیع‌های آماری با تکیه گاه مثبت که در قابلیت اعتماد با آن ها سرو کار داریم عبارتند از توزیع نمایی، توزیع وایبل، توزیع لگ نرمال و توزیع گاما. در ادامه این بخش نگاهی اجمالی به این مدل ها و دلایل استفاده از آن ها می اندازیم.

- توزیع نمایی: تابع چگالی و تابع خطر این توزیع به ترتیب عبارتند از

$$f(x) = \lambda e^{-\lambda x}, \qquad x > ۰, \lambda > ۰$$
$$h(x) = \lambda.$$

از مهمترین خاصیت های توزیع نمایی، خاصیت بی حافظه بودن آن است. با این که به دلیل سادگی کار، توزیع نمایی توزیع محبوبی است، نرخ خطر ثابت این توزیع محدودیت های بسیاری را در کارهای عملی در صنعت و بهداشت تحمیل می‌کند.

- توزیع وایبل: توزیع وایبل اولین بار توسط وایبل[۱] (۱۹۳۹)، فیزیکدان سوئدی معرفی شد که از آن برای نمایش توزیع شکاف مقاومت مصالح استفاده کرد. در توزیع وایبل با چگالی و نرخ خطر زیر، $\lambda > ۰$ پارامتر مقیاس و $\alpha > ۰$ پارامتر شکل است.

[۱] Weibull

$$f(x) = \alpha\lambda(\lambda x)^{\alpha-۱} e^{(-\lambda x)^\alpha}, \qquad x > ۰, \alpha, \lambda > ۰.$$
$$h(x) = \alpha\lambda(\lambda x)^{\alpha-۱}$$

یکی از خواص مهم توزیع وایبل، انعطاف پذیری تابع نرخ خطر آن است.

- توزیع لگ نرمال: متغیر تصادفی X را دارای توزیع لگ نرمال گوییم هرگاه لگاریتم طبیعی آن دارای توزیع نرمال باشد. این توزیع به دلیل رابطه ای که با توزیع نرمال دارد مورد استفاده زیاد قرار می گیرد. اما از آنجا که تابع نرخ خطر این توزیع برای طول عمر های زیاد، کوچک می شود در بسیاری از موارد نامناسب عمل می کند چرا که این خاصیت تابع نرخ خطر، در عمل خیلی کم اتفاق می افتد.

- توزیع گاما: اگر X دارای توزیع گاما با تابع چگالی زیر باشد

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-۱} e^{-\lambda x} \qquad x > ۰, \alpha, \lambda > ۰. \tag{۳.۱}$$

این توزیع دارای مشخصاتی شبیه به توزیع وایبل است، اما متأسفانه تابع نرخ خطر آن به صورت صریحی در دسترس نیست. وقتی که α، پارامتر شکل توزیع، به سمت بینهایت میل می کند، این توزیع به توزیع نرمال میل می کند.

از آنجایی که داده های قابلیت اعتماد معمولاً به صورت سانسور شده هستند در بخش بعد به معرفی انواع داده های سانسور شده می پردازیم.

۲.۱ سانسور

سانسور یک پدیده متداول در آزمون‌های طول عمر و مطالعات قابلیت اعتماد است. یک آزمایشگر ممکن است قادر نباشد، اطلاعات کاملی از زمان شکست همه واحد های مورد آزمایش به دست آورد. برای مثال افراد در یک آزمایش بالینی ممکن است که از ادامه آزمایش منصرف شوند و یا به علت کمبود بودجه آزمایش متوقف شود. در آزمایش‌های صنعتی نیز ممکن است واحد های آزمایش به منظور صرفه جویی

در زمان و هزینه، قبل از شکست و به صورت از پیش تعیین شده از آزمایش کنار گذاشته شوند. در چنین شرایطی داده هایی که تحلیل گر به دست می آورد، داده های سانسور شده هستند.

۱.۲.۱ سانسور از راست

یک آزمایش روی تعدادی از واحدها، ممکن است قبل از اینکه تمام واحدها شکست بخورند، به پایان برسد. داده های به دست آمده از چنین آزمایشی به داده های از راست سانسور شده موسومند. فرض کنید $X_1, X_2, ..., X_n$ طول عمر n واحد باشند که در یک آزمایش قرار گرفته اند، در این صورت داده های از راست سانسور شده به شکل زیر مشاهده می‌شوند

$$Y_i = \begin{cases} X_i & X_i \leq X_c \\ & \\ X_c & X_i > X_c \end{cases} , i = 1, ..., n$$

که در آن X_c نشانگر زمان سانسور است. در واقع اگر متغیر مورد مطالعه بسیار بزرگ باشد و نخواهیم آن را به طور کامل مشاهده کنیم از این سانسور استفاده می‌کنیم.

۲.۲.۱ سانسور از چپ

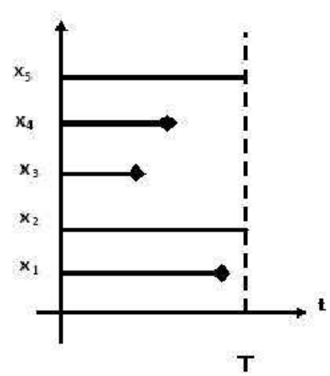
در این نوع سانسور، بعضی از واحد ها ممکن است قبل از زمان معین، شکست خورده باشند. برای چنین مشاهداتی می‌دانیم که پیشامد مطلوب قبل از زمان سانسور (X_c) اتفاق افتاده است، اما زمان دقیق آن معلوم نیست. به عبارت دیگر زمان دقیق بقای طول عمر X مربوط به یک فرد معین، معلوم خواهد بود اگر $X > X_c$ باشد. اگر $X_1, X_2, ..., X_n$ طول عمر واحد هایی باشند که در آزمایش قرار می‌گیرند، آنگاه داده های طرح سانسور از چپ به صورت زیرمشاهده می شوند

$$Y_i = \begin{cases} X_c & X_i < X_c \\ & \\ X_i & X_i \geq X_c \end{cases} , i = 1, ..., n$$

برای مثال در مراکز یادگیری کودکان، اغلب پیشامد مطلوب روی آزمون بچه ها برای تعیین زمان یادگیری کودک برای انجام کارهای مشخص متمرکز می شود، فرض کنید که سن یادگیری کودک، زمان تا رخداد پیشامد یادگیری باشد. بعضی از کودکانی که آزمون را شروع می‌کنند قبلاً کار مورد نظر را یاد گرفته اند. چنین زمان های پیشامدی از چپ سانسور شده اند.

۳.۲.۱ سانسور نوع I (سانسور زمان)

در برخی شرایط به علت وجود هزینه یا محدودیت زمانی نمی توانیم منتظر باشیم تا همه داده ها ثبت شوند و در زمانی دلخواه آزمایش را قطع می‌کنیم. فرض کنید n مؤلفه در زمان صفر شروع به کار می‌کنند و تا زمان معلوم T آزمایش ادامه پیدا می‌کند و در زمان T آزمایش خاتمه می یابد. در این زمان امکان دارد که بعضی از مؤلفه ها هنوز مشغول کار باشند، بنابراین طول عمر این قطعات ثبت نمی شود. داده هایی که به این ترتیب به دست می آیند یک نمونه سانسور شده نوع I را می سازند. این طرح سانسور از نوع سانسور راست است؛ زیرا مشاهدات بزرگتر کنار گذاشته شده اند که در این طرح سانسور تعداد مشاهدات یک متغیر تصادفی است. به عنوان مثال اگر ۵ لامپ را روشن نموده و زمان سوختن لامپ ها را ثبت کنیم، با توجه به شکل ۱.۱ طول عمر لامپ های شماره ۲ و ۵ ثبت نشده است و تنها اطلاعی که درباره طول عمر این لامپ ها داریم این است که طول عمر آن ها حداقل T بوده است.



شکل ۱.۱: نمودار طرح سانسور نوع I برای طول عمر لامپ ها

۴.۲.۱ سانسور نوع II (سانسور شکست)

در سانسور نوع I زمان کل آزمایش ثابت و در دست پژوهشگر است (T)ولی تعداد قطعاتی که در فاصله زمانی $[۰, T]$ از کار می افتند یک متغیر تصادفی است. با احتمال $(۱ - S(t))^n$ هیچ قطعه ای در فاصله $[۰, T]$ نمی سوزد و این امر باعث می شود تا دقت ما در برآورد پارامتر مجهول کاهش یابد. برای رفع این عیب از نوع دیگری از سانسور به نام سانسور نوع II استفاده می‌شود. درسانسور نوع II همه قطعات در زمان $t = ۰$ تحت آزمایش قرار می‌گیرند (شروع به کارمی‌کنند) و تا زمان خرابی r امین قطعه آزمایش ادامه خواهد داشت. اگر آزمایش را در زمان از کار افتادن r امین واحد، $۱ \leq r \leq n$، پایان دهیم نمونه سانسور شده نوع II به دست می آید. در این حالت r ثابت است. در حالی که زمان پایان آزمایش یعنی X_r، تصادفی است. این طرح سانسور نیز از نوع سانسور راست است؛ زیرا مقادیر بزرگ سانسور شده اند.

۳.۱ سانسور های فزاینده

دو نوع از متداول ترین طرح‌های سانسور یعنی سانسور های نوع I و II را در بخش قبل معرفی کردیم . یک نقطه ضعف این دو طرح سانسور آن است که امکان کنار گذاشتن واحدها در زمان‌هایی دیگر، غیر از زمان پایان آزمایش (یعنی زمان از پیش تعیین شده T یا زمان شکست r ام) را ندارند. در حالی که گاهی لازم است واحد ها را در مراحل مختلف آزمایش کنار بگذاریم. مخصوصاً هنگامی که داده‌های مورد آزمایش به سختی به دست می‌آیند و یا پرهزینه هستند. به همین دلیل سانسورهای کلی تری به نام سانسورهای فزاینده مطرح می شوند. هرد[۲] (۱۹۵۶) برای اولین بار مسأله سانسور فزاینده را مطرح نمود. این سانسور را معمولاً سانسور چندگانه یا سانسور چند مرحله ای نیز می‌نامند.علاقه مندان می توانند برای جزئیات بیشتر در مورد سانسور فزاینده به کارهای انجام شده توسط کوهن [۳] (۱۹۶۳،۱۹۶۶،۱۹۷۵،۱۹۷۶)، آگاروالا و بالاکریشنان[۴] (۱۹۹۶،۱۹۹۸)، بالاکریشنان و آگاروالا (۲۰۰۰) و بالاکریشنان و همکاران (۲۰۰۴) و بصیری (۱۳۸۹) و اسدی (۱۳۹۰) مراجعه نمایند. دراین بخش به معرفی انواع سانسورهای فزاینده از راست می پردازیم.

---

[۲] Herd

[۳] Cohen

[۴] Aggarwala and Balakrishnan

۱.۳.۱ سانسور فزاینده نوع اول از راست

در سانسور فزاینده نوع اول از راست، n واحد وارد آزمایش طول عمر می شوند. فرض می‌شود که آزمایشگر m زمان سانسور از پیش تعیین شده $T_m, ..., T_۲, T_۱$ را در نظر می‌گیرد و در زمان T_i، $i = ۱, ..., m - ۱$، تعداد R_i واحد سالم (که مقدار آن از قبل مشخص است) را به تصادف از آزمایش حذف (سانسور) می‌کند. البته به شرط این که R_i واحد سالم در زمان T_i، وجود داشته باشد. به عبارت دیگر واحد های سانسور شده در زمان T_i برابر

$$min\left(R_i, T_i \text{ تعداد واحدهای سالم در زمان } \right)$$

است. این آزمایش می تواند در هر یک از زمان های $T_{m-۱}, ..., T_۲, T_۱$ خاتمه یابد یا نهایتاً در زمان T_m که همه واحد های باقیمانده سانسور می شوند، پایان می‌یابد. علاوه بر زمان های سانسور T_i، زمان های شکست $X_k, ..., X_۲, X_۱$ را هم خواهیم داشت. بنابراین در این سانسور زمان های شکست و زمان های سانسور یکسان نیستند.

۲.۳.۱ سانسور فزاینده نوع دوم از راست

در طرح سانسور فزاینده نوع دوم از راست، n واحد در زمان صفر در یک آزمون طول عمر قرار می‌گیرند. بلافاصله پس از مشاهده اولین شکست، $R_۱$ واحد سالم از بین $n - ۱$ واحد دیگر به تصادف از آزمایش خارج می‌شوند. با مشاهده شکست دوم، $R_۲$ واحد سالم از بین $(n - R_۱ - ۲)$ واحد موجود سانسور می‌شوند. این روند ادامه می‌یابد تا این که در زمان شکست m ام همه $R_m = n - R_۱ - R_۲ - ... - R_{m-۱} - m$ واحد باقیمانده از آزمایش کنار گذاشته می شوند.

در این طرح سانسور مقادیر $R_m, ..., R_۱$ (و درنتیجه m) از قبل مشخص و ثابت می‌باشند و زمان های سانسور همگی تصادفی هستند. نتیجه آزمایش m مقدار مرتب شده است که با توجه به نوع سانسور، آنها را آماره های مرتب سانسور فزاینده نوع دوم از راست می نامند.

در حالت خاص هنگامی که $R_۱ = R_۲ = ... = R_{m-۱} = ۰$ و $R_m = n - m$، این طرح سانسور به طرح سانسور نوع II تبدیل می شود که در آن تنها m آماره مرتب مشاهده می شود.

همچنین اگر $R_۱ = R_۲ = ... = R_m = ۰$ ، آنگاه n آماره مرتب بدون سانسور(نمونه کامل) حاصل خواهد شد. بنابراین آماره های مرتب حالت خاصی از آماره های مرتب سانسور فزاینده نوع دوم از راست هستند.

۴.۱ ترتیب های تصادفی

مرتب سازی توزیع‌ها، مخصوصاً در بین توزیع‌های طول عمر، نقش مهمی در علوم آماری بازی می‌کند. ترتیب‌های تصادفی و نامساوی‌ها در طی چهل سال اخیر مورد توجه و استفاده فراوان قرار گرفته اند و در برخی از زمینه های آمار و احتمال ازجمله نظریه قابلیت اعتماد، نظریه صف و آنالیز بقا و . . . و همچنین در اقتصاد، بیمه و علم مدیریت کاربرد دارند. محققان بسیاری در زمینه ترتیب تصادفی کار کرده اند که می توان از لهمن[۵] (۱۹۵۵)، مارشال و اولکین[۶] (۱۹۷۹) و شیکد و شانتیکومار[۷] (۲۰۰۷) و ... نام برد. در ادامه تعریف های مورد نیاز فصل‌های بعدی این رساله، بیان می شوند.
فرض کنید X و Y دو متغیر تصادفی به ترتیب با توابع توزیع $F(x)$ و $G(y)$، توابع چگالی $f(x)$ و $g(y)$ و توابع بقا $\bar{F}(x) = ۱ - F(x)$ و $\bar{G}(y) = ۱ - G(y)$ هستند.

تعریف ۱.۴.۱. *متغیر تصادفی X را در ترتیب تصادفی نسبت درستنمایی کوچکتر از Y گوییم (و آن را با نماد $X \leq_{lr} Y$ نشان می دهیم) اگر $\frac{g(x)}{f(x)}$ نسبت به x صعودی باشد.*

تعریف ۲.۴.۱. *متغیر تصادفی X را در ترتیب تصادفی نرخ خطر کوچکتر از Y گوییم (و آن را با نماد $X \leq_{hr} Y$ نشان می دهیم) اگر $h_F(x) \geq h_G(x), \ \forall x \in R^+$ یا به طور معادل اگر $\frac{\bar{F}(x)}{\bar{G}(x)}$ تابعی نزولی از x باشد که h_F و h_G به ترتیب توابع نرخ خطر متغیر های تصادفی X و Y هستند.*

تعریف ۳.۴.۱. *متغیر تصادفی X را در ترتیب تصادفی معمولی (یا در توزیع) کوچکتر از Y گوییم (و آن را با نماد $X \leq_{st} Y$ نشان می دهیم) اگر $\bar{F}(x) \leq \bar{G}(x), \ \forall x \in (-\infty, \infty)$ باشد*

تعریف ۴.۴.۱. *متغیر تصادفی X را در ترتیب تصادفی پراکندگی کوچکتر از Y گوییم (و آن را با نماد $X \leq_{disp} Y$ نشان می دهیم) اگر و تنها اگر*

$$F^{-۱}(v) - F^{-۱}(u) \leq G^{-۱}(v) - G^{-۱}(u) \qquad ۰ < u \leq v < ۱$$

یا به طور معادل $G^{-۱}(u) - F^{-۱}(u)$ تابعی صعودی برای هر $u \in (۰, ۱)$ باشد که $F^{-۱}$ و $G^{-۱}$ به ترتیب توابع معکوس از راست پیوسته F و G هستند.

---

[۵] Lehmann

[۶] Marshall and Olkin

[۷] Shaked and Shanthikumar

تعریف ۱.۴.۵. *متغیر تصادفی X را در ترتیب تصادفی میانگین باقیمانده عمر کوچکتر از Y گوییم (و آن را با نماد $X \leq_{mrl} Y$ نشان می دهیم) اگر برای هر $m(t) \leq l(t),\ t \in R$ باشد که m و l به ترتیب توابع میانگین باقیمانده طول عمر متغیر های تصادفی X و Y هستند.*

تعریف ۱.۴.۶. *متغیر تصادفی X را در ترتیب تصادفی تبدیل محدب کوچکتر از Y گوییم (وآن را با نماد $X \leq_c Y$ نشان می دهیم) اگر $G^{-1}(F)$ تابعی محدب باشد.*

تعریف ۱.۴.۷. *متغیر تصادفی X را در ترتیب تصادفی ستاره کوچکتر از Y گوییم (وآن را با نماد $X \leq_* Y$ نشان می دهیم) اگر $\frac{G^{-1}(F(x))}{x}$ نسبت به $x > 0$ صعودی باشد و یابه طور معادل اگر و تنها اگر $\frac{G^{-1}(u)}{F^{-1}(u)}$ نسبت به $u \in (0, 1)$ صعودی باشد.*

نتیجه ۱.۴.۱. *اگر $X \leq_{lr} Y$ ، آنگاه $X \leq_{hr} Y$، اگر $X \leq_{hr} Y$ ، آنگاه $X \leq_{st} Y$. همچنین اگر $X \leq_{hr} Y$ ، آنگاه $X \leq_{mrl} Y$،اگر $X \leq_c Y$، آنگاه $X \leq_* Y.$ (شیکد و شانتیکومار، ۲۰۰۷).*

۱.۵ خواص مجانبی برآوردگرهای درستنمایی ماکسیمم

روش درستنمایی ماکسیمم ML یکی از متداول ترین و عمومی‌ترین روش‌ها در بین کاربران آمار است. به برآوردگرهایی که از این روش حاصل می شوند، برآوردگرهای درستنمایی ماکسیمم گوییم. این برآوردگرها نسبت به سایر برآوردگرها (مانند برآوردگرهای گشتاوری، برآوردگرهای خطی و . . .) ویژگی های جالبی دارند که به برخی از آن ها اشاره می‌کنیم.
فرض کنید جامعه مورد بررسی دارای تابع چگالی احتمال (یا تابع جرم احتمال) $f(x;\theta)$ باشد که به بردار پارامتر مجهول $\theta = (\theta_1, \theta_2, ..., \theta_k)$ بستگی دارد. براساس یک مجموعه داده که به روشی دلخواه جمع آوری شده است، برآورد درستنمایی ماکسیمم θ برداری است که تابع درستنمایی مربوطه را ماکسیمم کند. شکل تابع درستنمایی علاوه بر توزیع جامعه به روش جمع آوری داده ها نیز بستگی دارد. اگر نمونه کامل در اختیار باشد، این تابع برابر حاصلضرب توابع چگالی احتمال به ازای همه مقادیر نمونه است. اما درحالتی که داده ها سانسور شده باشند، تعیین تابع درستنمایی کمی پیچیده است.
شاید بتوان برخی از برجسته ترین ویژگی‌های برآوردگرهای ML رابه صورت زیر خلاصه کرد: