دانشگاه مازندران

دانشکده علوم ریاضی

پایان نامه دوره کارشناسی ارشد

رشته آمار گرایش آمار ریاضی

عنوان:

توزیع‌های لجستیک تعمیم یافته

استاد راهنما:

دکتر اکبر اصغرزاده

استاد مشاور:

دکتر احمد پوردرویش

نگارش :

لیلا اسماعیلی

بهمن ماه ۱۳۸۹

بسم الله الرحمن الرحیم

تقدیم به:

خانواده‌ی عزیزم

قدردانی

بارالها تو را سپاس که فضلت را کران نیست و شکرت را زبان نیست.

اتمام این پایان نامه حاصل تلاش، راهنمایی و مساعدت عزیزانی است که خود را قدردان زحمات آنان می‌دانم. بر خود لازم می‌دانم مراتب سپاس و تشکر خود را از کلیه اساتید و دانشمندانی که در دوران تحصیلی‌ام، مرا از انوار بی‌کران دانش خویش بهره‌مند ساخته‌اند، ابراز نمایم. از استاد محترم و گرانقدر جناب آقای دکتر اصغرزاده که در طول مدت نگارش پایان‌نامه از هیچ کمکی دریغ نکردند و دلسوزانه و مهربانانه و با تلاش خستگی ناپذیر مرا راهنمایی فرمودند بی‌نهایت سپاسگزارم، همچنین از استاد گرامی، آقای دکتر پوردرویش به پاس قبول زحمت مشاورت این پایان نامه سپاسگزارم. از داوران محترم، جناب آقای دکتر صادقپور و سرکار خانم دکتر محمدپور به پاس داوری این پایان‌نامه و رفع نواقص و کاستی‌های آن سپاسگزارم و سلامتی و توفیق روزافزون عزیزان را از خداوند متعال خواستارم.

از خانواده عزیزم که از کودکی، شور دانستن و لذت کشف و جستجو را در من بیدار کردند، استقامت در تلاش را بر من آموختن و تمام این سالها با فراهم کردن آرامش فکری و آسایش روحی، بسیاری از دشواریها را بر من آسان نمودند با تمام وجود قدردانم.

چکیــده

توزیع لجستیک، در مدلهای رشد و در نوع خاصی از رگرسیون بنام رگرسیون لجستیک کاربرد دارد. همچنین در مدلهای طول عمر نیز از این توزیع استفاده می‌شود. شکل توزیع لجستیک خیلی شبیه شکل توزیع نرمال می‌باشد با این تفاوت که در مرکز، توزیع نرمال قله‌ای تر و دم‌های نازک‌تری نسبت به توزیع لجستیک دارد. در مقالات، فرم‌های مختلفی برای تعمیم توزیع لجستیک در نظر گرفته شده است. در این پایان‌نامه، دو تعمیم مهم از توزیع لجستیک بنام‌های توزیع لجستیک چوله و توزیع لجستیک تعمیم یافته نوع اول معرفی و خواص ریاضی آنها مورد مطالعه قرار می‌گیرد. توزیع‌های لجستیک چوله و لجستیک تعمیم یافته نوع اول به ترتیب از ایده‌ی آزالینی (۱۹۸۵) و ایده خانواده توزیع‌های با تابع نرخ خطر معکوس متناسب بدست می آیند. برخی دیگر از تعمیم‌های توزیع لجستیک نیز در این پایان‌نامه بحث خواهد شد.

کلمات کلیدی: برآورد درستنمایی ماکزیم توزیع چوله، توزیع‌لجستیک، شبیه سازی مونت کارلو، رگرسیون لجستیک، گشتاورها، آنتروپی.

فهرست مندرجات

۱ مقدمه و اهداف پژوهش ۱۰

۱.۱ توزیع لجستیک ۱۱
۱.۱.۱ گشتاورها و دیگر اندازه ها ۱۲
۲.۱.۱ ارتباط با دیگر توزیع‌ها ۱۴
۳.۱.۱ تابع نرخ خطر و تابع نرخ خطر معکوس توزیع لجستیک ۱۴
۴.۱.۱ شبیه سازی از توزیع لجستیک ۱۶
۵.۱.۱ کاربردها ۱۶

۲.۱ تعمیم‌های توزیع لجستیک ۱۸

۳.۱ اهداف پژوهش ۱۸

۲ توزیع لجستیک چوله ۲۰

۱.۲ مقدمه ۲۱

۲.۲ ویژگیهای توزیع لجستیک چوله ۲۲

۳.۲ نمایش تابع چگالی و تابع توزیع برحسب سریها ۲۵

۴.۲ تابع مشخصه و تابع مولد گشتاور ۳۱

۵.۲ گشتاورها ۳۲

۶.۲ برآوردها ۳۸
۱.۶.۲ برآوردهای درستنمایی ماکسیمم ۳۹
۲.۶.۲ برآوردهای گشتاوری ۴۱
۳.۶.۲ ماتریس اطلاع فیشر ۴۲

۷.۲ شبیه سازی ۴۳

۸.۲ کاربردها ۴۴

۳ توزیع لجستیک تعمیم یافته‌ی نوع اوّل ۴۶

۱.۳ خانواده توزیع‌های با نرخ خطر معکوس متناسب ۴۸

۲.۳ توزیع لجستیک تعمیم یافته‌ی نوع اول ۴۸
۱.۲.۳ تابع چگالی و تابع توزیع ۴۹
۲.۲.۳ تابع مولد گشتاور ۵۰
۳.۲.۳ ارتباط با دیگر توزیع‌ها ۵۴

۳.۳ برآوردها ۵۵
۱.۳.۳ برآوردهای درستنمایی ماکسیمم α و β با فرض معلوم بودن m ۵۵
۲.۳.۳ برآورد درستنمایی ماکسیمم α با فرض معلوم بودن m و β ۵۸
۳.۳.۳ برآورد درستنمایی ماکسیمم β با فرض معلوم بودن m و α ۵۹
۴.۳.۳ برآوردهای گشتاوری α و β با فرض معلوم بودن m ۶۰
۵.۳.۳ برآورد گشتاوری α با فرض معلوم بودن m و β ۶۱
۶.۳.۳ برآورد گشتاوری β با فرض معلوم بودن α و m ۶۱

۴.۳ توزیع مقادیر غایی ۶۱

۴ تعمیم‌های دیگر توزیع لجستیک ۶۴

۱.۴ مقدمه ۶۵

۲.۴ توزیع لجستیک تعمیم یافته‌ی نوع دوّم ۶۵
۱.۲.۴ ویژگیها ۶۷
۲.۲.۴ ارتباط با دیگر توزیع‌ها ۶۸

۳.۴ توزیع لجستیک تعمیم یافته‌ی نوع سوم ۶۹

۴.۴ توزیع لجستیک تعمیم یافته‌ی نوع I گسترش یافته ۷۰

۵.۴ تعمیم نادراجا و کاتز ۷۲

۵ برآوردهای درستنمایی تقریبی پارامترهای مکان و مقیاس توزیع لجستیک چوله ۷۶

۱.۵ برآوردهای درستنمایی ماکسیمم ۷۷

۲.۵ برآوردهای درستنمایی تقریبی ۷۸

۳.۵ اطلاع فیشر مشاهده شده ۸۲

۴.۵ نتایج شبیه سازی ۸۳

۵.۵ احتمالات پوشش ۸۶

۶.۵ مثال عددی ۹۰

۶ **پیوست** ۹۲

۱.۶ تغییرات میانگین ۹۳

۲.۶ آنتروپی ۹۴

۱.۲.۶ آنتروپی رنی ۹۴

۲.۲.۶ آنتروپی شانون ۹۶

۳.۶ آماره‌های مرتب ۹۷

۴.۶ توزیع‌های مجانبی M_n و m_n ۱۰۰

۵.۶ اثبات روابط ۱۰۱

۶.۶ برنامه‌های کامپیوتری ۱۰۳

کتاب‌نامه ۱۲۰

واژه‌نامه ۱۲۴

فصل ۱

مقدمه و اهداف پژوهش

۱.۱ توزیع لجستیک

اولین بار ورهولتس [1] (۱۸۳۸ و ۱۸۴۵)[۳۸, ۳۹] تابع لجستیک را به عنوان منحنی رشد مورد استفاده قرار داد. استفاده از تابع لجستیک به عنوان یک منحنی رشد بر مبنای معادله دیفرانسیل زیر می‌باشد

$$\frac{dF(x)}{dx} = c[F(x) - A][B - F(x)], \tag{۱.۱}$$

که درآن c و A و B ثابت‌هایی هستند که $c > ۰$ و $B > A$.

یک متغیر تصادفی X، دارای توزیع لجستیک می‌باشد اگر تابع توزیع و تابع چگالی آن به ترتیب به صورت زیر باشند.

$$G(x; m, \beta) = \frac{۱}{۱ + exp(-\frac{x-m}{\beta})}, \qquad -\infty < x < \infty, \tag{۱.۲}$$

$$g(x; m, \beta) = \frac{exp(-\frac{x-m}{\beta})}{\beta\{۱ + exp(-\frac{x-m}{\beta})\}^{۲}}, \qquad -\infty < x < \infty, \tag{۱.۳}$$

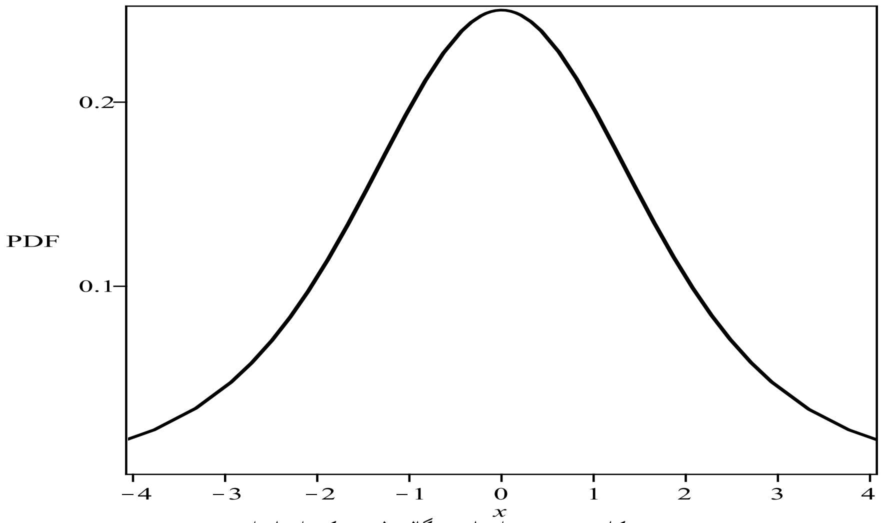
که $-\infty < m < \infty$ و $\beta > ۰$ می‌باشد، که در آن m و β به ترتیب پارامترهای مکان و مقیاس هستند. توزیع داده شده بالا حول پارامتر مکان m، متقارن می‌باشد. توزیع لجستیک را با $L(m, \beta)$ نمایش می‌دهند. توزیع لجستیک با $m = ۰$ و $\beta = ۱$ را توزیع لجستیک استاندارد نامیده و آن را با $L(۰, ۱)$ نمایش می‌دهیم. لذا تابع توزیع و تابع چگالی لجستیک استاندارد عبارتند از:

$$G(x) = \frac{۱}{۱ + \exp(-x)}, \qquad -\infty < x < \infty, \tag{۱.۴}$$

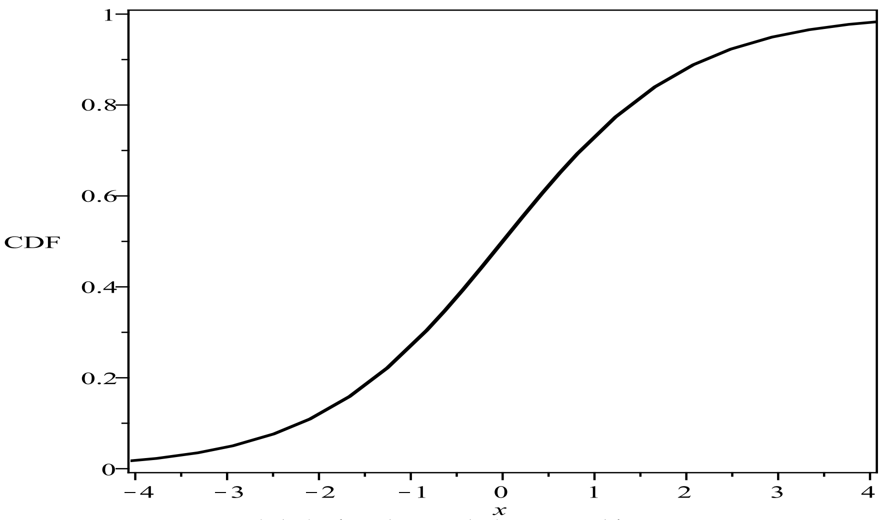
$$g(x) = \frac{\exp(-x)}{(۱ + \exp(-x))^{۲}}, \qquad -\infty < x < \infty. \tag{۱.۵}$$

شکل‌های (۱.۱) و (۱.۲)، تابع چگالی و تابع توزیع لجستیک استاندارد را نشان می‌دهد.

[1] Verhults



شکل ۱.۱: نمودار تابع چگالی لجستیک استاندارد



شکل ۱.۲: نمودار تابع توزیع لجستیک استاندارد

۱.۱.۱ گشتاورها و دیگر اندازه ها

اگر $X \sim L(0,1)$ باشد آنگاه تابع مولد گشتاورX عبارت است از:

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \exp(tx) \frac{\exp(-x)}{(1+\exp(-x))^2} dx, \\ &= \int_{-\infty}^{\infty} \frac{\exp(-x(1-t))}{(1+\exp(-x))^2} dx. \end{aligned}$$

با تغییر متغیر $u = \frac{۱}{۱+exp(-x)}$ داریم.

$$\begin{aligned} M_X(t) &= \int_{۰}^{۱} u^t(۱-u)^{-t}du, \\ &= B(t+۱, -t+۱), \\ &= \frac{\Gamma(t+۱)\Gamma(-t+۱)}{\Gamma(۲)}, \\ &= \Gamma(۱+t)\Gamma(۱-t), \qquad -۱ < t < ۱. \end{aligned}$$

برای محاسبه تابع مولد گشتاور توزیع لجستیک با پارامترهای m و β بصورت زیر اقدام می‌نماییم.

اگر $Y \sim L(m,\beta)$ باشد آنگاه با علم به اینکه $X = \frac{Y-m}{\beta}$ دارای توزیع لجستیک استاندارد می‌باشد داریم:

$$M_Y(t) = E(e^{t(m+\beta X)})) = E(e^{tm})E(e^{t\beta X}) = e^{tm}\Gamma(۱-t\beta).\Gamma(۱+t\beta).$$

برای بدست آوردن گشتاورهای توزیع لجستیک، از مشتق مراتب مختلف تابع مولد گشتاور استفاده می‌کنیم و گشتاورهای مختلف را بدست می آوریم.

تابع مولد کمولانت[۲] این توزیع برای حالت استاندارد چنین می‌باشد.

$$Q_X(t) = \ln M_X(t) = \ln\Gamma(۱-t) + \ln\Gamma(۱+t), \qquad -۱ < t < ۱. \qquad (۱.۷)$$

با استفاده از تابع مولد کمولانت و با توجه به اینکه می‌دانیم $\mu_r = E(X-\mu)^r = \frac{d^r}{dt^r}Q_X(t)|_{t=۰}$، بنابراین چهار گشتاور اول برای توزیع $L(۰,۱)$ به صورت زیر بدست می آیند.

$$\mu_۱ = ۰, \qquad \mu_۲ = \frac{\pi^۲}{۳}, \qquad ,\mu_۳ = ۰, \qquad ,\mu_۴ = \frac{۷\pi^۴}{۱۵}.$$

[۲] Cumulant generating function

۲.۱.۱ ارتباط با دیگر توزیع‌ها

– توزیع لجستیک، آمیخته‌ی توزیع نمایی و توزیع مقادیر غایی (گامبل) می‌باشد. برای روشن شدن مطلب، توجه کنید که تابع چگالی توزیع لجستیک استاندارد در (۱.۵) یعنی $g(x)$ را می‌توان بصورت زیر نوشت

$$\begin{aligned} g(x) &= \int_0^\infty \Big\{\text{چگالی توزیع گامبل}\Big\} \times \Big\{\text{چگالی توزیع نمایی}\Big\}\, d\alpha \\ &= \int_0^\infty \underbrace{\alpha e^{-x} e^{-\alpha e^{-x}}} \times \underbrace{e^{-\alpha}}\, d\alpha. \end{aligned}$$

– اگر Y_1 و Y_2 متغیرهای تصادفی نمایی مستقل با میانگین‌های λ_1 و λ_2 باشد آنگاه به کمک یک ژاکوبین ساده می‌توان نشان داد که متغیر تصادفی $Z = \ln \frac{Y_1}{Y_2}$ دارای توزیع لجستیک با پارامتر مکان $\ln \frac{\lambda_1}{\lambda_2}$ و پارامتر مقیاس ۱ با تابع چگالی زیر می‌باشد، یعنی

$$f(z) = \frac{\lambda_1 \lambda_2 e^z}{(\lambda_1 + \lambda_2 e^z)^2}.$$

– اگر X_1 و X_2 دو متغیر تصادفی مستقل و همتوزیع با توزیع گامبل استاندارد با تابع چگالی احتمال

$$\exp(-x)\exp(-\exp(-x))$$

باشند. در اینصورت می‌توان نشان داد که توزیع $X_1 - X_2$ توزیع لجستیک استاندارد می‌باشد.

۳.۱.۱ تابع نرخ خطر و تابع نرخ خطر معکوس توزیع لجستیک

تابع نرخ خطر، تابع نرخ خطر معکوس و تابع قابلیت اطمینان یک متغیر تصادفی X، با تابع توزیع $G_X(x)$ و تابع چگالی $g_X(x)$، به ترتیب با $h_X(.)$، $r_X(.)$ و $R_X(.)$ نمایش داده می‌شوند و به صورت زیر تعریف می‌باشند.

$$\begin{aligned} h_X(.) &= \frac{g(x)}{1 - G(x)}, \\ r_X(.) &= \frac{g(x)}{G(x)}, \\ R_X(.) &= 1 - G_X(.). \end{aligned}$$

اگر $X \sim L(۰, ۱)$ باشد در اینصورت تابع خطر X عبارت است از:

$$\begin{aligned} h_X(.) &= \frac{g(x)}{۱ - G(x)}, \\ &= \frac{\frac{e^{-x}}{(۱+e^{-x})^۲}}{۱ - \frac{۱}{۱+e^{-x}}}, \\ &= \frac{۱}{۱ + e^{-x}}, \\ &= G(x). \end{aligned}$$

لذا در توزیع لجستیک استاندارد، تابع نرخ خطر برابر تابع توزیع می‌باشد. بنابراین در توزیع لجستیک، تابع نرخ خطر یک تابع صعودی است. تابع نرخ خطر معکوس توزیع لجستیک استاندارد می‌شود

$$\begin{aligned} r(x) &= \frac{g(x)}{G(x)}, \\ &= \frac{g(-x)}{۱ - G(-x)}, \\ &= h(-x). \end{aligned}$$

لذا تابع نرخ خطر معکوس توزیع لجستیک در نقطه‌ی x با تابع نرخ خطر همین توزیع در نقطه‌ی $x-$ برابر می‌باشد. به همین ترتیب می‌توان نشان داد که به ازای تمام مقادیر x رابطه‌های زیر برقرار است.

$$\begin{aligned} r(x) &= R(x), \\ g(x) &= h(x).r(x), \\ h(x) + r(x) &= ۱, \\ \frac{h(x)}{r(x)} &= e^x. \end{aligned}$$

۴.۱.۱ شبیه سازی از توزیع لجستیک

به کمک قضیه تبدیل انتگرال احتمال، اگر متغیر تصادفی پیوسته X دارای تابع توزیع $G(x)$ باشد در اینصورت $U = G(X)$ دارای توزیع $U(0,1)$ می‌باشد. برای توزیع لجستیک چون $G(x) = \frac{1}{1+e^{-x}}$ است، لذا داریم.

$$U = \frac{1}{1+e^{-X}} \Longleftrightarrow X = \ln\left(\frac{U}{1-U}\right)$$

لذا برای تولید نمونه‌ی n تایی $x_1, x_2, \cdots, x_n$ از توزیع $L(0,1)$، ابتدا یک نمونه $u_1, u_2, \cdots, u_n$ از $U(0,1)$ تولید کرده و سپس قرار می‌دهیم.

$$x_i = \ln\frac{u_i}{1-u_i} \qquad i = 1, 2, \cdots, n$$

پس شبیه سازی از توزیع لجستیک به سه صورت زیر انجام می‌شود.

- اگر U دارای توزیع یکنواخت باشد یعنی $U \sim U(0,1)$ آنگاه $\ln(\frac{U}{1-U})$ دارای $L(0,1)$ می‌باشد.
- اگر $G(.)$ هر تابع توزیع مطلقاً پیوسته باشد آنگاه $\ln(\frac{G(.)}{1-G(.)})$ نیز دارای توزیع لجستیک استاندارد می‌باشد.
- با توجه به اینکه برای هر توزیع مطلقاً پیوسته با تابع توزیع $G(.)$، تابع نرخ خطر $h(.)$ و تابع نرخ خطر معکوس $r(.)$، $\ln(\frac{G(.)}{1-G(.)}) = \frac{h(.)}{r(.)}$، لذا $\ln(\frac{h(.)}{r(.)})$ دارای توزیع لجستیک استاندارد می‌باشد.

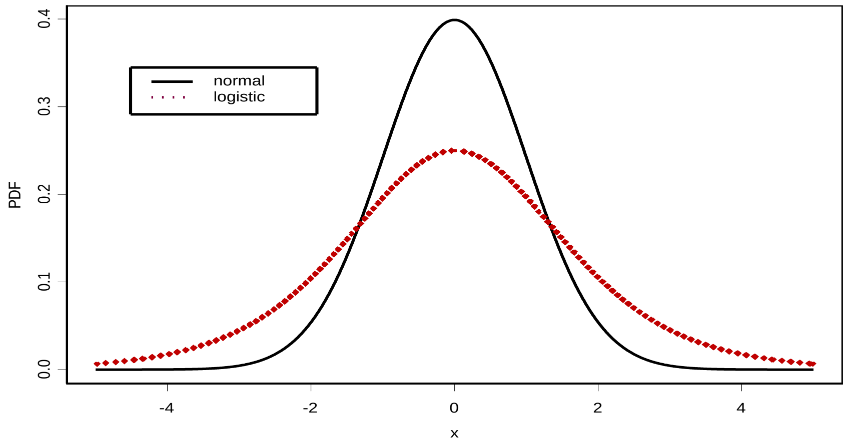
۵.۱.۱ کاربردها

تابع لجستیک، اولین بار توسط دانشمندان و محققان بسیاری به عنوان منحنی رشد مورد استفاده قرار گرفت. استفاده از منحنی رشد برای اهداف جمعیت شناسی اقتصادی در اواخر قرن نوزدهم خیلی مرسوم شد، در طی سالها کاربردهای بسیار دیگری برای این توزیع یافته شد.

این توزیع در تجزیه و تحلیل داده‌های بقا، در مطالعه‌ی توزیع درآمد یا ثروت، در بدست آوردن رابطه‌ی بین متغیرها و به عنوان جایگزینی برای توزیع نرمال (به دلیل سادگی آن نسبت به توزیع نرمال) مورد استفاده قرار می‌گیرد .

این توزیع در حالت استاندارد، حول صفر متقارن می‌باشد و نقاط عطف آن $\pm\, 0/53$ می‌باشد. شکل توزیع لجستیک

بسیار به شکل توزیع نرمال شباهت دارد با این تفاوت که توزیع نرمال در مرکز، کشیده‌تر از توزیع لجستیک می‌باشد. توزیع لجستیک دمهای سنگین‌تری از توزیع نرمال دارد. به دلیل شباهت بسیار توزیع لجستیک به توزیع نرمال این توزیع به عنوان جایگزینی برای توزیع نرمال مورد استفاده قرار می‌گیرد. شکل (۱.۳) بیانگر این مسئله می‌باشد.



شکل ۱.۳: نمودار تابع چگالی لجستیک استاندارد و توزیع نرمال استاندارد

برای استفاده از توزیع لجستیک به عنوان جایگزینی برای توزیع نرمال از اصولی استفاده می‌شود از جمله اطلاع کولبک —لایبر[۳] یا آنتریی نسبی[۴] و یا استفاده از اختلاف تابع توزیع و بسیاری از روشهای دیگر.

اطلاع کولبک— لایبر (KL) به صورت زیر تعریف می شود.

$$
\begin{aligned}
k(g,f) &= E_g\left[\ln\frac{g(x)}{f(x)}\right], \\
&= \int \ln(\frac{g(x)}{f(x)})g(x)dx.
\end{aligned}
$$

این مقدار اندازه‌ای از توانایی نسبت درستنمایی (LR) برای متمایز کردن برضد $f(x)$ زمانی که $g(x)$ درست است، می‌باشد. با استفاده از این اصول، دامنه‌ای از x را که در آن اختلاف این توزیع‌ها بسیار کم هست را در نظر می‌گیرند. خاصیت دم‌های سنگین توزیع لجستیک می‌تواند برای تولید کردن توزیع‌هایی با دم‌های سنگین مورد استفاده قرار گیرد.

[۳] Kullback-Leibler
[۴] Relative entropy

۲.۱ تعمیم‌های توزیع لجستیک

در سالهای اَخیر، توزیع‌های چوله نقش مهمی را در تجزیه و تحلیل داده‌های آماری و همچنین در منابع آماری ایفا می‌کنند. به همین منظور برای توزیع لجستیک تعمیم‌هایی ایجاد شده است، که اکثر این تعمیم‌ها، چوله و توزیع لجستیک حالت خاصی از آنها می‌باشد، در این پایان‌نامه به بررسی پنج تعمیم از توزیع لجستیک می‌پردازیم، تعمیم اوّل با استفاده از ایده‌ی آزالینی می‌باشد. آزالینی[۵] و کاپیتانیو[۶] (۲۰۰۳) [۹] و آزالینی[۷] (۲۰۰۵) [۸] اصولی را برای چوله کردن توزیع‌های متقارن ارائه داده‌اند که با گسترش این اصول برای توزیع لجستیک می‌توانیم توزیع لجستیک چوله را بدست آوریم. بر این اساس تابع چگالی توزیع لجستیک چوله استاندارد به صورت زیر تعریف می‌شود

$$f(x) = 2g(x)G(\lambda x),$$

که در آن $g(.)$ و $G(.)$ به ترتیب تابع چگالی و تابع توزیع لجستیک استاندارد می‌باشند و $\lambda \in R$، پارامتر چولگی می‌باشد. این توزیع را با $SL(\lambda)$ نشان می‌دهند. با تغییر متغیر $Y = \beta X + m$ که در آن $X \sim SL(\lambda)$ می‌توان توزیع را به توزیع لجستیک چوله‌ی غیراستاندارد $SL(m, \beta, \lambda)$ تبدیل نمود که در آن m پارامتر مکان، β پارامتر مقیاس و λ پارامتر چولگی می‌باشد. تعمیم دوم، توزیع لجستیک تعمیم یافته‌ی نوع اوّل می‌باشد که در فصل سوّم به بررسی آن می‌پردازیم. به همین ترتیب توزیع لجستیک تعمیم یافته‌ی نوع دوّم و نوع سوّم و نوع معرفی شده به وسیله‌ی اولا پدا[۸] و نوع ارائه شده توسط نادراجا[۹] معرفی خواهند شد.

۳.۱ اهداف پژوهش

در این پایان‌نامه، فصل دوّم به طور مفصّل به بررسی ویژگیهای توزیع لجستیک چوله می‌پردازد که شامل برآورد پارامترها و ویژگیهای این توزیع می‌باشد. فصل سوّم، شامل ویژگیهای توزیع لجستیک تعمیم یافته‌ی نوع اوّل،

---
[۵] Azzalini
[۶] Capitanio
[۷] Azzalini
[۸] Olapeda
[۹] Nadrajah

مسئله‌ی برآوردهای درستنمایی ماکسیمم و برآوردهای گشتاوری پارامترهای این توزیع می‌باشد، علاوه برآن توزیع مقادیر غایی نیز محاسبه می‌شود. فصل چهارم به بررسی تعمیم‌های دیگری از توزیع لجستیک می‌پردازد. فصل پنجم با استفاده از روشهای درستنمایی ماکسیمم و روش درستنمایی ماکسیمم تقریبی به برآورد پارامترهای توزیع لجستیک چوله می‌پردازد.

فصل ۲

توزیع لجستیک چوله