# برآوردهای توانمند در مدل رگرسیون پواسون

سید امیررضا محمودی[۱]، روح‌الله روزگار [۲]

تاریخ دریافت: ۱۴۰۱/۰۷/۲۴
تاریخ پذیرش: ۱۴۰۲/۰۲/۲۲

**چکیده:**
بسیاری از روش‌های برآوردیابیِ رگرسیونی در مواجه با داده‌های پرت به‌شدت تحت تأثیر قرار می‌گیرند و خطاهای زیادی در برآوردهای حاصل از آن‌ها رخ می‌دهد. در سال‌های اخیر، برای حل این مشکل برآوردگرهایی توانمند توسعه‌یافته‌اند. برآوردگر حداقل واگرایی توان چگالی یک روش برآورد بر مبنای حداقل فاصله بین دو تابع چگالی است که این روش، در مواجه با موقعیت‌هایی که داده‌ها شامل تعدادی داده پرت هستند برآوردگرهایی توانمند ارائه می‌دهد. در این پژوهش، روش برآوردگر توانمند حداقل واگرایی توان چگالی را برای برآورد پارامترهای مدل رگرسیون پواسون ارائه می‌کنیم که می‌تواند برآوردگرهای توانمند با کمترین نقصان در کارایی تولید کند. همچنین عملکرد برآوردگرهای پیشنهادی را از طریق ارائه مثال واقعی موردبررسی قرار خواهیم داد.
**واژه‌های کلیدی:** برآوردگر حداقل واگرایی توان چگالی، پارامتر همسان‌ساز، توانمندی، داده پرت، رگرسیون پواسون ،مدل خطی تعمیم‌یافته.

## ۱ مقدمه

آمار علم ریاضی تفسیر و توضیح داده‌ها است و تحلیل‌های رگرسیونی یکی از مهم‌ترین ابزارهای کاربردی در علم آمار است. روش حداقل مربعات [۳] ($LS$) یکی از محبوب‌ترین برآوردگرهای رگرسیون است که در آثار گاوس و لژاندر معرفی شد (برای برخی از بحث‌های تاریخی به پلاکت (۱۹۷۲)، استیگلر (۱۹۸۱) و استیگلر (۱۹۸۶) مراجعه کنید). روش $LS$ که مربوط به بهینه‌سازی برازش با به حداقل رساندن مجموع مربعات باقیمانده است، به دلیل سهولت محاسبه آن پذیرفته‌شده است.

اما وجود مشاهدات پرت، هم در متغیرهای وابسته و هم در متغیرهای توضیحی، می‌تواند بر روی برآوردگرهای $LS$ تأثیر زیادی داشته باشد. در یک قرن گذشته حجم وسیعی از مقالات در مورد داده‌های پرت [۴] و خطرات آن‌ها برای روش‌های آماری کلاسیک ارائه‌شده است. برای حل این مشکل، تکنیک‌های جدید آماری مانند روش‌های توانمند توسعه‌یافته‌اند. در واقع توانمندی [۵] به معنای عدم حساسیت در برابر برخی انحرافات از مدل واقعی است. روش‌های توانمند در کارهای پیرسون، باکس و توکی (۱۹۶۰) و نظریه‌های رسمی‌تر در مورد توانمندی در دهه ۱۹۷۰ توسعه یافتند.

در مدل‌های رگرسیون، هدف روش‌های توانمند شناسایی نقاط پرت، نقاط بانفوذ [۶] و در نهایت توصیف بهترین برازش برای داده‌هاست. یکی از اولین گزینه‌های جایگزین برای روش پرکاربردِ حداقل مربعات ($LS$) ، به‌عنوان یک برآوردگر توانمند، توسط ادوورس (۱۸۸۷) پیشنهاد شد که برآوردگر حداقل انحراف مطلق [۷] ($LAD$) نامیده شد. نوآوری بعدی برآوردگرِ $M$ هوبر [۸] ($Huber$(۱۹۷۳)$Huber$(۱۹۸۳)) [۱۲] و [۱۳] بود. سپس به‌منظور محدود کردن تأثیر نقاط بانفوذ بالا، برآوردگرهای نوعِ $M$ تعمیم‌یافته به نام برآوردگرهای مالوز (۱۹۷۵) [۱۴] پیشنهاد شد. در این زمینه روسیو (۱۹۸۴) روش حداقل میانه مربع [۹] ($LMS$) را پیشنهاد کرد.

[۱] دانش‌آموخته کارشناسی ارشد رشته آمار، گروه ریاضی، دانشگاه یاسوج
[۲] هیئت‌علمی گروه ریاضی، دانشگاه یاسوج (نویسنده مسئول: roozegar@yu.ac.ir)

[3] Least Squares
[4] outliers
[5] Robustness
[6] leverage points
[7] Least Absolute Deviation
[8] Huber's M-estimator
[9] Least Median Squared

روش‌های توانمند برآورد پارامترهای مدل خطی تعمیم‌یافته در سال‌های اخیر موردتوجه برخی از پژوهشگران قرار گرفته است. در این مورد کانتونی و رانچتی (۲۰۰۱) روش توانمند برآورد شبه-درستنمایی هوبر [۱۰] $(HQ-LE)$ و حسینیان (۲۰۰۹) [۱۱] روش توانمند برآورد ماکسیمم درستنمایی وزنی [۱۱] $(WMLE)$ را برای برآورد پارامترهای مدل خطی تعمیم‌یافته ارائه داده‌اند. دوریو و ایسایا [۸] ازاین‌روش در برآورد توانمند مدل رگرسیون خطی استفاده کردند. یکی از حالت‌های خاص مدل‌های خطی تعمیم‌یافته مدل رگرسیون پواسون است. در این زمینه هال و شن (۲۰۱۰) [۱۰] ، ابونازل و صابر (۲۰۲۰) [۱] ، ابونازل و داوود (۲۰۲۲) [۲] ، امین و همکاران (۲۰۲۱) [۳] ، چن و همکاران (۲۰۱۸) [۷] و شان تسو (۲۰۰۵) [۱۷] سعی کردند از روش‌های توانمندی برای برآورد پارامترهای مدل رگرسیون پواسون استفاده کنند؛ اما نکته حائز اهمیت این است که توانمندی تنها معیار بهینه در برآورد پارامترها نمی‌باشد بلکه برآوردگری مناسب است که در عین توانمندی کارایی مناسبی نیز داشته باشد. در این مقاله سعی داریم با روش مبتنی بر حداقل واگرایی توان چگالی، برآوردگرهایی توانمند و درعین‌حال کارا برای پارامترهای مدل رگرسیون پواسون ارائه کنیم. خانواده فواصل مبتنی بر چگالی که به نام واگرایی توان چگالی معروف است برای اولین بار توسط باسو و همکاران [۴] در سال ۱۹۹۸ معرفی شد. برای آشنایی بیشتر با این خانواده فواصل مبتنی بر چگالی می‌توان به کارهای باسو [۴] ، سیمپسون [۱۵] و بران [۵] مراجعه کرد. این خانواده به پارامتر همسان‌ساز و نامنفی $\alpha$ وابسته است که این پارامتر تعادل بین کارایی [۱۲] و توانمندی را در این خانواده برقرار می‌کند. لذا اهمیت و مزیت مدلی که در این مقاله معرفی می‌شود نسبت به مدل‌های توانمند اشاره‌شده در بالا، کارایی مناسب در عین توانمندی است.

در ادامه تعریف جامعی از روش برآوردگر حداقل واگرایی توان چگالی [۱۳] $(MDPDE)$ ارائه می‌دهیم، سپس ازاین‌روش در برآورد پارامترهای مدل رگرسیون پواسون [۱۴] استفاده می‌کنیم و نشان می‌دهیم که این روش دارای خصوصیات مناسبی از جمله توانمندی و کارایی است. سپس عملکرد $MDPDE$های پیشنهادی را در مدل رگرسیون پواسون با اعمال آن بر روی یک مجموعه داده واقعی بررسی خواهیم کرد و نشان خواهیم داد که برآوردهای حاصل ازاین‌روش در مقابل مشاهدات پرت و دارای نفوذ، توانمند بوده و کارایی قابل قبولی دارند. چندین نمودار و جدول نیز برای فهم و درک بهتر مفاهیم ارائه‌شده است.

## ۲ تعاریف، فرضیات و قضایا

مواردی پیش می‌آید که ارتباط بین متغیر پاسخ و متغیر توضیحی خطی نیست و یا متغیرهای پاسخ پیوسته نمی‌باشند و یا ممکن است که توزیع متغیر پاسخ نرمال نباشد. در چنین مواقعی مدل خطی تعمیم‌یافته [۱۵] $(GLM)$ می‌تواند به‌عنوان یک ابزار کلی و بسیار مهم راهگشا باشد. $GLM$ به آزمایشگر اجازه می‌دهد تا متغیرهای پاسخ با هر توزیعی از میان خانواده توزیع‌های نمایی را با هر ترکیب خطی از متغیرهای توضیحی مدل‌سازی کند.

مدل‌های خطی تعمیم‌یافته همان‌طور که از اسم آن‌ها پیداست تعمیمی از مدل‌های خطی می‌باشند. مدل‌های خطی در مواقعی که متغیر پاسخ ($Y_i$ها) پیوسته یا نزدیک به توزیع نرمال باشند، به فرم زیر خواهند بود:

$$E(Y_i)=\mu_i=\boldsymbol{x}_i^T\boldsymbol{\beta}\ ; \qquad Y_i \sim N(\mu_i,\sigma^2) \qquad (۱)$$

مدل‌های خطی به فرم بالا که در آن متغیرهای تصادفی $Y_i$ مستقل هستند، اساس اکثر تحلیل‌های داده‌های پیوسته هستند. متغیرهای تصادفی $Y_i$ برای موضوعات مختلف، ممکن است مقادیر مورد انتظار $\mu_i$ متفاوتی داشته باشد. بردار $\boldsymbol{x}_i^T$ نشان‌دهنده ردیف $i$-ام ماتریس طرح $\boldsymbol{X}$ است.

پیشرفت‌ها در تئوری آماری و نرم‌افزار رایانه‌ای به ما امکان می‌دهد از روش‌هایی مشابه روش‌های توسعه‌یافته برای مدل‌های خطی در موقعیت‌های کلی‌تر زیر استفاده کنیم:

۱- متغیرهای پاسخ دارای توزیعی غیر از توزیع نرمال باشند و حتی ممکن است به‌جای پیوسته، گروه‌بندی باشند.

۲- رابطه بین متغیر پاسخ و متغیرهای توضیحی به فرم خطی ساده در ۱ نباشد.

در مواقعی که متغیر پاسخ ($Y_i$) جزء خانواده نمایی باشد با مدل‌های غیرخطی سروکار داریم که بین متغیر پاسخ و توضیحی رابطه مستقیم خطی وجود ندارد. بلکه پس از اعمال تابع پیوندی مانند $g$ روی میانگین متغیر پاسخ، رابطه خطی به‌صورت زیر به دست می‌آید:

$$g(\mu_i)=\boldsymbol{x}_i^T\boldsymbol{\beta}$$

[10] Huber Quasi-Likelihood Estimate
[11] Weighted Maximum Likelihood Estimate
[12] Efficiency
[13] Minimum Density Power Divergence Estimator
[14] Poisson Regression
[15] Generalized Linear Model

تابع $g$، تابع پیوند نامیده می‌شود که یک تابع ریاضی ساده است.

## ۱.۲ مدل‌های خطی تعمیم‌یافته

وحدت بسیاری از روش‌های آماری توسط نلدر و ودربورن (۱۹۷۲) با استفاده از ایده یک مدل خطی تعمیم‌یافته نشان داده شد. این مدل بر اساس مجموعه‌ای از متغیرهای تصادفی مستقل $Y_1,\dots,Y_N$ تعریف‌شده که هرکدام از $Y_i$ ها عضو خانواده توزیع‌های نمایی است. ویژگی‌های مدل به‌صورت زیر است:

۱-توزیع هرکدام از $Y_i$ ها به فرم کانونی بوده و تنها به پارامتر $\theta_i$ بستگی دارد،بنابراین

$$f(y_i;\theta_i)=exp[y_ib_i(\theta_i)+c_i(\theta_i)+d_i(y_i)].$$

۲-توزیع تمام $Y_i$ ها یکسان بوده و تابع چگالی احتمال توأم از $Y_1,\dots,Y_N$ به‌صورت زیر است

$$\begin{aligned}f(y_1,\dots,y_N;\theta_1,\dots,\theta_N)&=\prod_{i=1}^{N} exp[y_ib(\theta_i)+c(\theta_i)+d(y_i)]\\&=exp\left[\sum_{i=1}^{N}y_ib(\theta_i)+\sum_{i=1}^{N}c(\theta_i)+\sum_{i=1}^{N}d(y_i)\right]\end{aligned} \quad (۲)$$

فرض کنید $E(Y_i)=\mu_i$ ،به‌طوری‌که $\mu_i$ تابعی از $\theta_i$ است. برای یک مدل خطی تعمیم‌یافته تبدیلی از $\mu_i$ مانند زیر وجود دارد

$$g(\mu_i)=\boldsymbol{x}_i^T\boldsymbol{\beta}.$$

در این رابطه برای $i=1,\dots,N$ داریم:

الف. تابع پیوند $g$ یک تابع یکنوا و دیفرانسیل پذیر است.

ب. بردار $\boldsymbol{x}_i$ یک بردار $p\times 1$ از متغیرهای توضیحی است.

ج. $\boldsymbol{\beta}$ بردار $p\times 1$ از پارامترهاست. بردار $\boldsymbol{x}_i^T$ ردیف $i$ ام از ماتریس طرح $\boldsymbol{X}$ است.

د. متغیرهای پاسخ $Y_1,\dots,Y_N$ توزیع یکسانی از خانواده نمایی دارند.

در مدل‌های خطی تعمیم‌یافته $(GLM)$ با انتخاب چگالی‌های گوناگون از خانواده نمایی به‌جای $f$ و تابع پیوندهای گوناگونِ $g$ ، مدل‌های رگرسیون متفاوتی به دست می‌آید . به‌عنوان‌مثال اگر $f$ چگالی نرمال و $g$ تابع پیوند همانی انتخاب شوند ، مدل خطی تعمیم‌یافته به مدل رگرسیون خطی نرمال معمولی تنزل می‌یابد و یا اگر $f$ چگالی پواسون و $g$ تابع پیوند لگاریتمی $g(\mu)=\log(\mu)$ انتخاب شوند ، حالت رگرسیون پواسون به دست می‌آید که در مدل‌سازی داده‌های شمارش‌پذیر استفاده می‌شود. اگر $f$ چگالی دوجمله‌ای و تابع پیوند لجیت [16] $g(\mu)=\log(\mu/(1-\mu))$ یا تابع پیوند پروبیت [17] $g(\mu)=\Phi^{-1}(\mu)$ انتخاب شوند آنگاه به ترتیب به ترتیب مدل‌های رگرسیون لجستیک و پروبیت را تولید می‌کنند به‌طوری‌که برای مدل‌سازی متغیرهای پاسخ دو دویی استفاده می‌شوند.

یک حالت بسیار پرکاربرد از مدل‌های خطی تعمیم‌یافته رگرسیون پواسون می‌باشد که آن را به‌طور مفصل موردبررسی قرار می‌دهیم. روش معمول در برآورد پارامترهای مدل رگرسیون خطی تعمیم‌یافته روش درستنمایی ماکسیمم $(MLE)$ و روش حداقل مربعات $(LS)$ می‌باشد؛ اما این روش‌های کلاسیک به وجود داده پرت در مجموعه داده‌ها حساس بوده و برآوردهای ارائه‌شده توسط این روش‌ها تحت تأثیر داده‌های پرت به‌شدت از مقدار صحیح خود فاصله می‌گیرند که در اصطلاح می‌گوییم این روش‌ها توانمند نیستند؛ بنابراین ارائه روش‌های توانمند بسیار ضرورت دارد. در این مقاله از روش برآوردگر حداقل واگرایی توان چگالی $(MDPDE)$ در برآورد پارامترهای مدل خطی تعمیم‌یافته به‌طور عام و مدل رگرسیون پواسون به‌طور خاص استفاده می‌کنیم.

## ۲.۲ اندازه واگرایی توان چگالی

به‌طورکلی در استنباط بر اساس حداقل فاصله، برآورد پارامتر مورد نظر خود را به‌وسیله مینیمم کردن یک اندازه مناسب از فاصله بین داده‌ها و مدل پارامتری فرض شده (پیشنهادشده) به دست می‌آوریم.

**تعریف ۱.۲.** *(باسو و همکاران) [۴] و گوش و باسو [۹]) اگر $X_1,X_2,\dots,X_n$ نمونه‌ای تصادفی به حجم $n$ از تابع چگالی واقعی $g$ باشد که $g$ ناشناخته باشد و بخواهیم این تابع چگالی را با تابع چگالی$f$ برآورد کنیم، آنگاه اندازه واگرایی توان چگالی$(DPD)$ [18] بین دو تابع توزیع $g$ و $f$ که با نماد $d_\alpha(g,f)$ نشان داده می‌شود به‌صورت زیر تعریف می‌شود.*

*برای حالت $\alpha>0$:*

$$\begin{aligned}d_\alpha(g,f)=\int\{f^{1+\alpha}(x;\boldsymbol{\theta})-(\frac{1+\alpha}{\alpha})f^\alpha(x;\boldsymbol{\theta})g(x)\\+\frac{1}{\alpha}g^{1+\alpha}(x)\}d(x)\end{aligned} \quad (۳)$$

*و برای حالت $\alpha=0$:*

$$d_0(g,f)=\lim_{\alpha\to 0}d_\alpha(g,f)=\int g(x)\log(\frac{g(x)}{f_{\boldsymbol{\theta}}(x)})d(x) \quad (۴)$$

[16] Logit link function
[17] Probit link function
[18] Density Power Divergence

*در عمل، $g$ توزیع داده‌ها و $f$ توزیع مدل است (که وابسته به پارامتر مجهول است). این اندازه (تابع) تنها به همان پارامتر همسان‌ساز $\alpha$ بستگی دارد. تابع $d_\alpha(g,f)$ یک اندازه است. بدین معنا که این عبارت همواره غیر منفی است و زمانی صفر می‌شود که $g \equiv f$ باشد.*

فرض کنید که بخواهیم بردار پارامتری $\boldsymbol{\theta}$ را برآورد کنیم. همچنین فرض کنید

$$\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq R^p\}$$

خانواده پارامتری از توابع و $G$ تابع توزیع متناظر با تابع چگالی $g$ باشد و $g \in \mathcal{G}$ که $\mathcal{G}$ کلاس تمام توزیع‌ها با تکیه‌گاه یکسان است. حال اگر برآوردگر مورد نظر یا همان برآوردگر واگرایی توان چگالی بردار پارامتری $\boldsymbol{\theta}$ را به‌وسیله $T_\alpha(G)$ نشان دهیم، آنگاه خواهیم داشت:

$$d_\alpha(g, f_{T_\alpha(G)}) = \min_{\boldsymbol{\theta} \subseteq \boldsymbol{\Theta}} d_\alpha(g, f_{\boldsymbol{\theta}})$$

مفهوم این رابطه این است که اگر $g$ تابع چگالی صحیح ما باشد، آنگاه از بین فضای پارامتری مربوط به $\boldsymbol{\theta}$، توسط برآوردگر حداقل واگرایی توان چگالی آن بردار عددی را انتخاب می‌کنیم که فاصله بین تابع چگالی صحیح ($g$) و تابع چگالی برآورد شده توسط برآوردگر حداقل واگرایی توان چگالی ($f_{T_\alpha(G)}$) کمترین فاصله را داشته باشد.

## ۳.۲ برآوردگر حداقل واگرایی توان چگالی برای مشاهدات مستقل و هم‌توزیع

در قضیه‌های زیر به محاسبه برآورد حداقل واگرایی توان چگالی، زمانی که از یک نمونه مستقل و هم‌توزیع استفاده کنیم می‌پردازیم.

**قضیهٔ ۲.۲.** *(گوش و باسو (۲۰۱۳) [۹]) خانواده توابع چگالی: $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ را در نظر بگیریم و همچنین اگر فرض کنیم $X_1, \ldots, X_n$ یک نمونه تصادفی از توزیع $G$ باشد و $d_\alpha(g,f)$ به صورتی که در قبل توضیح داده‌ایم باشد. در این صورت با استفاده از روش گوش و باسو (۲۰۱۳) [۹] برآوردگر حداقل واگرایی توان چگالی بردار پارامتری $\boldsymbol{\theta}$ از مینیمم‌کردن عبارت زیر حاصل می‌شود:*

$$\begin{aligned} H_\alpha(\boldsymbol{\theta}) &= \int f_{\boldsymbol{\theta}}^{1+\alpha}(x)d(x) - (1+\tfrac{1}{\alpha})\ \tfrac{1}{n}\textstyle\sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\alpha}(X_i) \\ &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\ V_{n,\boldsymbol{\theta}}(X_i) \end{aligned}$$

*به‌طوری‌که* $V_{n,\boldsymbol{\theta}}(X) = \int f_{\boldsymbol{\theta}}^{1+\alpha}(x)d(x) - (1+\frac{1}{\alpha}) f_{\boldsymbol{\theta}}^{\alpha}(X)$ .

**قضیهٔ ۳.۲.** *(گوش و باسو [۹]) اگر $u_{\boldsymbol{\theta}}(x) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(x)$ آنگاه تحت شرایط مشتق‌پذیری مدل، می‌توانیم برآورد حداقل واگرایی توان چگالی بردار پارامتری $\boldsymbol{\theta}$ برای یک نمونه مستقل و هم‌توزیع را از رابطه زیر محاسبه کنیم:*

$$\frac{1}{n}\ \sum_{i=1}^{n} u_{\boldsymbol{\theta}}(X_i) f_{\boldsymbol{\theta}}^{\alpha}(X_i)\ -\ \int u_{\boldsymbol{\theta}}(x_i) f_{\boldsymbol{\theta}}^{1+\alpha}(x) d(x) = 0.$$

**تذکر ۴.۲.** *معادله برآوردیابی که در قضیه ۳.۲ آمده، در حقیقت یک مدل وزنی از معادله حداکثر درستنمایی است، به این صورت که برای تمامی $\alpha > 0$ این معادله به مشاهدات پرت وزن پایین‌تری نسبت به مدل و به مشاهداتی که به مدل مفروض نزدیک‌تر باشند وزن بالاتری اختصاص می‌دهد. در مقابل در حالتی که $\alpha = 0$ باشد، این معادله به تمام مشاهدات، حتی مشاهداتی که آن‌ها را به‌عنوان داده پرت می‌شناسیم وزنی برابر یک اختصاص می‌دهد؛ بنابراین با انتخاب $\alpha$ نزدیک به صفر، می‌توانیم وزن تمامی مشاهدات را نزدیک به یک بگیریم که این کار باعث بهبود کارایی مجانبی برآوردگر در مقایسه با روش حداقل فاصله می‌شود. از طرف دیگر هر چه $\alpha$ به یک نزدیک‌تر باشد، توانمندی این برآوردگر در مواجه با داده‌های پرت افزایش پیدا می‌کند. نکته‌های گفته‌شده دقیقاً به برقراری توازن بین کارایی و توانمندی برآوردگر توسط پارامتر همسان‌ساز $\alpha$ اشاره دارد.*

## ۴.۲ برآوردگر حداقل واگرایی توان چگالی برای مشاهدات مستقل و غیرهم‌توزیع

گاهی با مواردی روبه‌رو می‌شویم که مشاهدات ما اگرچه از هم مستقل هستند ولی توزیع‌های متفاوتی دارند.

فرض کنید که داده‌های مشاهده‌شده ما به‌صورت $Y_1, \ldots, Y_n$ باشند که از هم مستقل بوده اما برای هر $i = 1, \ldots, n$ ، $Y_i \sim g_i$ باشد؛ یعنی هر یک از نقاط $Y_i$ ، متعلق به توابع چگالی جداگانه به‌صورت $g_1, \ldots, g_n$ است؛ بنابراین برای پیش‌بینی هر $g_i$ از مدل $f_i$ و $i = 1, \ldots, n$ متعلق به خانواده توابع چگالی $F_{i,\boldsymbol{\theta}} = f_i(.;\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}, i = 1, 2, \ldots, n$ استفاده می‌کنیم. این توابع اگرچه ممکن است توزیع‌های متفاوتی داشته باشند ولی همگی آن‌ها در بردار پارامتری $\boldsymbol{\theta}$ مشترک هستند. هدف اصلی ما برآورد کردن بردار پارامتری $\boldsymbol{\theta}$ با استفاده از به حداقل رساندن اختلاف توان چگالی بین داده‌ها و مدل در نظر گرفته‌شده است.

### ۱.۴.۲ روش باسو وقتی مشاهدات مستقل و غیرهم‌توزیع باشند

در این حالت چگالی مدل $f_i$ و توزیع واقعی $g_i$ برای هر نقطه متفاوت است، بنابراین برای برآورد پارامتر مورد نظر ما باید اختلاف بین داده‌ها و مدل را به‌طور جداگانه برای هر نقطه محاسبه کرده و سپس سعی کنیم میانگین تمامی این فواصل را مینیمم کنیم. به‌عبارت‌دیگر اگر فاصله بین چگالی واقعی $g_i$ و مدل $f_i$ را با $d_\alpha(g_i, f_i(y;\boldsymbol{\theta}))$ نشان دهیم داریم:

$$d_\alpha(g_i, f_i(y;\boldsymbol{\theta})) = \int \{f_i^{1+\alpha}(y;\boldsymbol{\theta}) - (1+\frac{1}{\alpha}) f_i^{\alpha}(y;\boldsymbol{\theta}) g_i + \frac{1}{\alpha} g_i^{1+\alpha}$$

که عبارت سومی مستقل از پارامتر $\boldsymbol{\theta}$ بوده و در فرایند مینیمم سازی تأثیری ندارد. چون در بیشتر موارد چگالی واقعی $g_i$ ناشناخته است، در محاسبات خود از برآورد آن استفاده می‌کنیم. اگر $d_\alpha(\hat{g}_i, f_i(.;\boldsymbol{\theta}))$ را به‌عنوان فاصله بین $i$ اُمین $\hat{g}_i$ و چگالی مدل $f_i$ تعریف کنیم، می‌توانیم برای به حداقل رساندن تمام فواصل بالا به ازای $i = 1, \dots, n$ با به حداقل رساندن

$$\frac{1}{n} \sum_{i=1}^{n} d_\alpha(\hat{g}_i, f_i(.;\boldsymbol{\theta})), \boldsymbol{\theta} \in \boldsymbol{\Theta}$$

به هدف خود برسیم؛ اما برای برآورد کردن چگالی واقعی داده‌ها معمولاً از چگالی تجربی استفاده می‌کنیم. در چگالی تجربی به هر نقطه احتمال $\frac{1}{n}$ اختصاص داده می‌شود و چون در این حالت از هر تابع چگالی واقعی $g_i$ ، تنها یک نقطه $Y_i$ در دسترس داریم بنابراین بهترین تخمین ممکن چگالی $g_i$ توزیع تباهیده شده در نقطه $Y_i$ است و داریم:

$$\begin{aligned} d_\alpha(\hat{g}_i, f_i(.;\boldsymbol{\theta})) &= \int f_i(y;\boldsymbol{\theta})^{1+\alpha} dy \\ &\quad -(1+\tfrac{1}{\alpha}) f_i(Y_i;\boldsymbol{\theta})^{\alpha} + K \end{aligned} \tag{۵}$$

که $K$ یک ثابت و مستقل از پارامتر $\boldsymbol{\theta}$ است؛ بنابراین برای برآورد کردن پارامتر $\boldsymbol{\theta}$ از روش حداقل واگرایی توان چگالی به روش باسو برای مشاهدات مستقل و غیرهم‌توزیع باید تابع هدف زیر را به حداقل برسانیم.

$$\begin{aligned} H_n(\boldsymbol{\theta}) &= \tfrac{1}{n} \sum_{i=1}^{n} \{\int f_i(y;\boldsymbol{\theta})^{1+\alpha} dy - (1+\tfrac{1}{\alpha}) f_i(Y_i;\boldsymbol{\theta})^{\alpha}\} \\ &= \tfrac{1}{n} \sum_{i=1}^{n} V_i(Y_i;\boldsymbol{\theta}) \end{aligned} \tag{۶}$$

به‌طوری‌که $V_i(Y_i, \boldsymbol{\theta}) = \int f_i^{1+\alpha}(y;\boldsymbol{\theta}) dy - (1+\frac{1}{\alpha}) f_i(Y_i;\boldsymbol{\theta})^{\alpha}$ است.

**نتیجه ۵.۲.** *اگر فرض کنیم $Y_1, \dots, Y_n$ یک نمونه تصادفی مستقل و غیرهم‌توزیع باشد به‌طوری‌که $Y_i \sim g_i$ و $i = 1, \dots, n$ ، آنگاه تحت شرایط مشتق‌پذیری مدل، برآوردگر حداقل واگرایی توان چگالی پارامتر $\boldsymbol{\theta}$ را به روش باسو می‌توان از عبارت زیر محاسبه کرد:*

$$\sum_{i=1}^{n} \{f_i(Y_i;\boldsymbol{\theta})^{\alpha} u_i(Y_i;\boldsymbol{\theta}) - \int f_i(y;\boldsymbol{\theta})^{1+\alpha} u_i(y;\boldsymbol{\theta}) dy\} = 0 \tag{۷}$$

*زمانی‌که* $u_i(y;\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f_i(y;\boldsymbol{\theta})$

**تذکر ۶.۲.** *اگر $\alpha \to 0$ بنابراین برای به دست آوردن برآورد پارامتر $\boldsymbol{\theta}$ در حالتی‌که مشاهدات مستقل و غیرهم‌توزیع باشند کافی است تابع هدف زیر را مینیمم کنیم:*

$$\frac{1}{n} \sum_{i=1}^{n} \{-\ln(f_i(Y_i;\boldsymbol{\theta}))\}.$$

*ملاحظه می‌کنید که عبارت به‌دست‌آمده همان برآوردگر ماکسیمم درستنمایی است؛ بنابراین نتیجه می‌گیریم برآوردگر حداقل واگرایی توان چگالی زمانی‌که $\alpha = 0$ باشد معادل با برآوردگر ماکسیمم درستنمایی است؛ بنابراین معادله برآوردی ۷ یک تعمیم ساده از برآوردگر ماکسیمم درستنمایی می‌باشد.*

**تذکر ۷.۲.** *به‌طورکلی برآوردگر حداقل واگرایی توان چگالی $T_\alpha(G_1, \dots, G_n)$ برای مشاهدات مستقل ولی غیرهمگن، آن برآوردگری است که در عبارت زیر صدق کند*

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} d_\alpha(g_i(.), f_i(.;T_\alpha(G_1, \dots, G_n))) \\ &= \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{n} \sum_{i=1}^{n} d_\alpha(g_i(.), f_i(.;\boldsymbol{\theta})) \end{aligned} \tag{۸}$$

*یعنی برآوردگر حداقل واگرایی توان چگالی آن مقداری از پارامتر $\boldsymbol{\theta}$ در فضای پارامتری $\Theta$ است که به ازای آن میانگین تمام فواصل بین $g_i$ و $f_i$ ها حداقل شود.*

## ۵.۲ برآوردگر حداقل واگرایی توان چگالی برای مدل‌های خطی تعمیم‌یافته

### ۱.۵.۲ تعریف مدل خطی تعمیم‌یافته:

گوییم $n$ متغیر مستقل $Y_1, \dots, Y_n$ از مدل خطی تعمیم‌یافته تبعیت می‌کند هرگاه تابع چگالی این متغیرها از فرم نمایی تعمیم‌یافته به‌صورت زیر پیروی کند

$$f(y_i;\theta_i,\phi) = exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi) \right\} \tag{۹}$$

وقتی‌که پارامتر کانونی $\theta_i$ (پارامتر مکان) وابسته به متغیرهای مستقل $X_i$ باشد و $\phi_i$ نیز به‌عنوان پارامتر مزاحم باشد. همچنین میانگین

متغیرهای $Y_i$ یعنی $\mu_i$ برابر باشد با $g(\mu_i) = \boldsymbol{\eta}_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$ به‌طوری‌که $g$ تابع پیوند و یک تابع یکنوا و مشتق‌پذیر باشد. پس ما مجموعه‌ای از مشاهدات مستقل و غیر هم‌توزیع $y_1, \ldots, y_n$ داریم که برای $y_i$ برای $i = 1, \ldots, n$ دارای چگالی

$$f_i(\ .\ ; (\boldsymbol{\beta}, \phi)) \ = \ f(y_i \ ; \ \theta_i \ , \ \phi) \tag{10}$$

است. هدف ما برآورد بردار پارامتری $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$ است؛ بنابراین بر اساس مشاهدات بالا و با استفاده از روش باسو کافی است در رابطه شماره ۶ به‌جای $f$ تابع چگالی ۱۰ و به‌جای بردارِ $\boldsymbol{\theta}$ پارامترهای $(\boldsymbol{\beta}, \phi)$ را قرار داده تا برای مشاهدات مستقل و غیر هم‌توزیع ، برآوردگر $MDPDE$ را به دست آوریم. ازاین‌رو برای مشاهدات مستقل و غیر هم‌توزیع ، برآوردگر حداقل واگرایی توان چگالی ($MDPDE$) مربوط به مدل خطی تعمیم‌یافته برای پارامترهای $(\boldsymbol{\beta}, \phi)$ با مینیمم سازیِ عبارت زیر به دست می‌آید:

$$H_n(\boldsymbol{\beta}, \phi) = \frac{1}{n} \sum_{i=1}^{n} V_i(Y_i; (\boldsymbol{\beta}, \phi)) \tag{11}$$

به‌طوری‌که

$$V_i(Y_i; (\boldsymbol{\beta}, \phi)) = \int {f_i}^{1+\alpha}(y; (\boldsymbol{\beta}, \phi))dy - (1 + \frac{1}{\alpha}){f_i}^{\alpha}(Y_i; (\boldsymbol{\beta}, \phi)).$$

در برآورد حداقل $DPD$ پیشنهادی، به‌طور همزمان پارامترهای $\boldsymbol{\beta}$ و $\phi$ را فقط با به حداقل رساندن $H_n(\boldsymbol{\beta}, \phi)$ با توجه به هر دو پارامتر به‌صورت توانمند برآورد می‌کنیم.

**نتیجه ۸.۲.** *با توجه به نتیجه ۵.۲ تحت شرایط مشتق‌پذیری مدل، با جایگذاری پارامترهای $(\boldsymbol{\beta}, \phi)$ به‌جای بردارِ $\boldsymbol{\theta}$ ، برآوردگر حداقل واگرایی توان چگالی پارامترهای $\boldsymbol{\beta}$ و $\phi$ را به روش باسو می‌توان از حل عبارت‌های زیر به دست آورد:*

$$\sum_{i=1}^{n} \nabla V_i(Y_i; (\boldsymbol{\beta}, \phi)) = 0, \tag{12}$$

*یا*

$$\sum_{i=1}^{n} \left[ \int u_i(y; (\boldsymbol{\beta}, \phi)) f_i(y; (\boldsymbol{\beta}, \phi))^{1+\alpha} dy - u_i(Y_i; (\boldsymbol{\beta}, \phi)) f_i(Y_i; (\boldsymbol{\beta}, \phi))^{\alpha} \right] = 0 \tag{13}$$

*به‌طوری‌که* $u_i(y; (\boldsymbol{\beta}, \phi)) = \nabla \log f_i(y; (\boldsymbol{\beta}, \phi))$

لازم است بدانیم که $\nabla$ مشتق را با توجه به $(\boldsymbol{\beta}, \phi)$ نشان می‌دهد. مشتقات جزئی $\nabla_{\boldsymbol{\beta}}$ و $\nabla_{\phi}$ را با استفاده از قانون زنجیره‌ای برای مشتق‌گیری به‌صورت زیر محاسبه می‌کنیم:

$$\nabla_{\boldsymbol{\beta}} \log(f_i(y_i; (\boldsymbol{\beta}, \phi))) = \frac{(y_i - \mu_i)}{Var(Y_i)g'(\mu_i)} x_i = K_{1i}(y_i; (\boldsymbol{\beta}, \phi)) x_i, \tag{14}$$

$$\nabla_{\phi} \log(f_i(y_i; (\boldsymbol{\beta}, \phi))) = -\frac{(y_i\theta_i - b(\theta_i))}{a^2(\phi)} a'(\phi) + \frac{\partial}{\partial \phi} c(y_i, \phi) = K_{2i}(y_i; (\boldsymbol{\beta}, \phi)) \tag{15}$$

به‌طوری‌که $K_{1i}$ و $K_{2i}$ توابع نشانگر هستند؛ بنابراین با جایگذاری روابط ۱۴ و ۱۵ در رابطه‌ی ۱۳ معادلات برآوردی به‌صورت زیر می‌شوند:

$$\sum_{i=1}^{n} x_i \left[ \int K_{1i}(y; (\boldsymbol{\beta}, \phi)) f_i(y; (\boldsymbol{\beta}, \phi))^{1+\alpha} dy - K_{1i}(y_i; (\boldsymbol{\beta}, \phi)) f_i(y_i; (\boldsymbol{\beta}, \phi))^{\alpha} \right] = 0 \tag{16}$$

$$\sum_{i=1}^{n} \left[ \int K_{2i}(y; (\boldsymbol{\beta}, \phi)) f_i(y; (\boldsymbol{\beta}, \phi))^{1+\alpha} dy - K_{2i}(y_i; (\boldsymbol{\beta}, \phi)) f_i(y_i; (\boldsymbol{\beta}, \phi))^{\alpha} \right] = 0 \tag{17}$$

اگر بخواهیم که پارامتر $\boldsymbol{\beta}$ را با ثابت در نظر گرفتن $\phi$ برآورد کنیم، کافی است تا فقط معادله برآوردی ۱۶ در نظر گرفته شود. به‌علاوه برای حالتی که $\alpha = 0$ باشد قسمت اول رابطه ۱۶ برابر صفر خواهد بود، یعنی داریم:

$$\begin{aligned}
&\int \frac{(y - \mu_i)}{Var(y)g'(\mu_i)} x_i f_i(y; (\boldsymbol{\beta}, \phi))^{1+\alpha} dy \\
&= \frac{x_i}{Var(y)g'(\mu_i)} \left[ \int (y - \mu_i) f_i(y; (\boldsymbol{\beta}, \phi))^{1+\alpha} dy \right] \\
&= \frac{x_i}{Var(y)g'(\mu_i)} \left[ \int y f_i(y; (\boldsymbol{\beta}, \phi)) dy - \mu_i \int f_i(y; (\boldsymbol{\beta}, \phi)) dy \right] \\
&= \frac{x_i}{Var(y)g'(\mu_i)} (E(Y) - \mu_i) \\
&= \frac{x_i}{Var(y)g'(\mu_i)} (\mu_i - \mu_i) = 0
\end{aligned} \tag{18}$$

و معادلات برآوردی برای $\boldsymbol{\beta}$ (با نادیده گرفتن $\phi$ ) به‌صورت زیر ساده می‌شود:

$$\sum_{i=1}^{n} \frac{(Y_i - \mu_i)}{Var(Y_i)g'(\mu_i)} x_i = 0 \tag{19}$$

برآوردگر حداقل واگرایی توان چگالی زمانی که $\alpha = 0$ باشد معادل با برآوردگر ماکسیمم درستنمایی است؛ بنابراین معادله برآوردی ۱۳ یک تعمیم ساده از برآوردگر ماکسیمم درستنمایی می‌باشد. همچنین اگر چگالی $f$ به‌صورت $\int f(y; (\theta_i, \phi))^{1+\alpha} dy$ باشد که از پارامتر مکانی $\theta_i$ مستقل است، مانند چگالی توزیع نرمال ، آنگاه داریم:

$$\int \frac{(y_i - \mu_i)}{Var(y_i)g'(\mu_i)} x_i f_i(y_i; (\boldsymbol{\beta}, \phi))^{1+\alpha} dy = 0$$

و بنابراین معادله برآوردی ۱۶ به‌صورت زیر ساده می‌شود:

$$\sum_{i=1}^{n} \frac{(Y_i-\mu_i)}{Var(Y_i)g'(\mu_i)} x_i f_i(Y_i;(\boldsymbol{\beta},\phi))^{\alpha} = 0. \qquad (20)$$

## ۶.۲ ویژگی‌های مجانبی برآوردگر حداقل واگرایی توان چگالی

برآوردگر حداقل واگرایی توان چگالی دارای برخی ویژگی‌های مجانبی مانند توزیع مجانبی، کارایی مجانبی و توانمندی مجانبی است که در ادامه به آن‌ها اشاره می‌کنیم. در ابتدا توزیع توأم مجانبی برآوردگر حداقل واگرایی توان چگالی $(\hat{\beta},\hat{\phi})$ پارامترهای $(\boldsymbol{\beta},\phi)$ که با حل معادلات برآوردی ۱۶ و ۱۷ به‌دست‌آمده را به دست خواهیم آورد. فرض می‌کنیم که توزیع واقعی مولد داده‌ها نیز به چگالی مدل با پارامترهای $(\boldsymbol{\beta}^g,\phi^g)$ تعلق دارد. برای $i=1,\ldots,n$ و $j,k=1,2$ عبارات زیر را تعریف می‌کنیم:

$$\gamma_{ji}=\gamma_{ji}^{1+\alpha}(\boldsymbol{\beta},\phi)=\int K_{ji}(y;(\boldsymbol{\beta},\phi))f_i(y;(\boldsymbol{\beta},\phi))^{1+\alpha}dy,$$

$$\gamma_{jki}=\gamma_{jki}^{1+\alpha}(\boldsymbol{\beta},\phi)=\int K_{ji}(y;(\boldsymbol{\beta},\phi))K_{ki}(y;(\boldsymbol{\beta},\phi)) f_i(y;(\boldsymbol{\beta},\phi))^{1+\alpha}dy$$

به‌طوری‌که:

$$\boldsymbol{N}_i^{1+\alpha}(\boldsymbol{\beta},\phi)=\int u_i(y;(\boldsymbol{\beta},\phi))f_i(y;(\boldsymbol{\beta},\phi))^{1+\alpha}dy=\begin{pmatrix}\gamma_{1i}\boldsymbol{x}_i\\ \gamma_{2i}\end{pmatrix},$$

$$\boldsymbol{M}_i^{1+\alpha}(\boldsymbol{\beta},\phi)=\int u_i(y;(\boldsymbol{\beta},\phi))u_i(y;(\boldsymbol{\beta},\phi))^T f_i(y;(\boldsymbol{\beta},\phi))^{1+\alpha}dy =\begin{pmatrix}\gamma_{11i}\boldsymbol{x}_i\boldsymbol{x}_i^T & \gamma_{12i}\boldsymbol{x}_i\\ \gamma_{12i}\boldsymbol{x}_i^T & \gamma_{22i}\end{pmatrix}.$$

اکنون برای $k=1,2$ و $X^T=[x_1,\ldots,x_n]$ ماتریس‌های $\Gamma_j^{(\alpha)}$ و $\Gamma_{jk}^{(\alpha)}$ و همچنین مقادیر $\Psi_n$ و $\Omega_n$ را به‌صورت زیر تعریف می‌کنیم:

$$\Gamma_j^{(\alpha)}=Diag(\gamma_{ji})_{i=1,\ldots,n} \quad , \quad \Gamma_{jk}^{(\alpha)}=Diag(\gamma_{jki})_{i=1,\ldots,n} \quad .$$

$$\Psi_n(\boldsymbol{\beta},\phi)=\frac{1}{n}\sum_{i=1}^{n}M_i^{1+\alpha}(\boldsymbol{\beta},\phi) \qquad (21)$$

$$\Omega_n(\boldsymbol{\beta},\phi)=\frac{1}{n}\sum_{i=1}^{n}\left[M_i^{1+2\alpha}(\boldsymbol{\beta},\phi)-N_i^{1+\alpha}(\boldsymbol{\beta},\phi)(N_i^{1+\alpha}(\boldsymbol{\beta},\phi))^T\right] \qquad (22)$$

**لم ۹.۲.** *اگر فرض کنیم $Y_1,\ldots,Y_n$ یک نمونه تصادفی مستقل و غیر هم توزیع باشد به‌طوری‌که $Y_i\sim g_i$ و $i=1,\ldots,n$ آنگاه، با توجه به تعاریف بالا، برای این نمونه مستقل و غیرهمگن مقادیر $\Psi_n$ و $\Omega_n$ به‌صورت زیر به دست می‌آید:*

$$\Psi_n(\boldsymbol{\beta},\phi)=\frac{1}{n}\sum_{i=1}^{n}M_i^{1+\alpha}(\boldsymbol{\beta},\phi) =\frac{1}{n}\begin{pmatrix}X^T\Gamma_{11}^{(\alpha)}X & X^T\Gamma_{12}^{(\alpha)}1\\ 1^T\Gamma_{12}^{(\alpha)}X & 1^T\Gamma_{22}^{(\alpha)}1\end{pmatrix}; \qquad (23)$$

$$\Omega_n(\boldsymbol{\beta},\phi)=\frac{1}{n}\sum_{i=1}^{n}\left[M_i^{1+2\alpha}(\boldsymbol{\beta},\phi)-N_i^{1+\alpha}(\boldsymbol{\beta},\phi)(N_i^{1+\alpha}(\boldsymbol{\beta},\phi))^T\right]$$
$$=\frac{1}{n}\begin{pmatrix}X^T[\Gamma_{11}^{(2\alpha)}-(\Gamma_1^{(\alpha)})^T\Gamma_1^{(\alpha)}]X & X^T[\Gamma_{12}^{(2\alpha)}-\Gamma_1^{(\alpha)}\Gamma_2^{(\alpha)}]1\\ 1^T[\Gamma_{12}^{(2\alpha)}-\Gamma_1^{(\alpha)}\Gamma_2^{(\alpha)}]X & 1^T[\Gamma_{22}^{(2\alpha)}-(\Gamma_2^{(\alpha)})^T\Gamma_2^{(\alpha)}]1\end{pmatrix}. \qquad (24)$$

*که در اینجا ۱ یک بردار $n$ بعدی با عناصر ۱ و به‌صورت $1^T=[1,\ldots,1]$ است.*

*اثبات.* مقادیر $\Psi_n$ و $\Omega_n$ با جایگذاری $\boldsymbol{M_i}^{1+\alpha}$ ها و $\boldsymbol{N_i}^{1+\alpha}$ ها با روابط و تعاریف متناظر آن‌ها و انجام محاسبات جبری ساده به دست می‌آیند. □

با توجه به مقدمات بالا و با استفاده از مفروضات ۷گانه که در گوش باسو (۲۰۱۳) [۹] ارائه‌شده، در زیر به خواص مجانبی برآوردگر حداقل واگرایی توان چگالی برای مشاهدات مستقل و غیرهمگن می‌پردازیم.

### ۱.۶.۲ خواص مجانبی برآوردگر حداقل واگرایی توان چگالی برای مشاهدات مستقل و غیرهم‌توزیع

**قضیهٔ ۱۰.۲.** *(گوش و باسو [۹] ) تحت فرض‌های ۱ تا ۷ گفته‌شده، نتایج زیر برقرار است:*

*۱) یک دنباله سازگار $\hat{\boldsymbol{\theta}}_n=(\hat{\boldsymbol{\beta}}_n,\hat{\phi}_n)$ از ریشه‌های معادله برآوردگر حداقل واگرایی توان چگالی (روابط ۱۶ و ۱۷) وجود دارد.*

*۲) توزیع مجانبی $\Omega_n^{-\frac{1}{2}}\Psi_n[\sqrt{n}((\hat{\boldsymbol{\beta}}_n,\hat{\phi}_n)-(\boldsymbol{\beta}^g,\phi^g))]$، توزیع نرمال $(p+1)$ بعدی با (بردار) میانگین صفر و ماتریس واریانس-کواریانس $I_{p+1}$ می‌باشد، $I_{p+1}$ همان ماتریس همانی $p+1$ بعدی است. به‌طوری‌که $\Omega_n=\Omega_n(\boldsymbol{\beta}^g,\phi^g)$ و $\Psi_n=\Psi_n(\boldsymbol{\beta}^g,\phi^g)$ است.*

از قضیه فوق چنین نتیجه می‌گیریم که معکوسِ ماتریس $\Psi_n^{-1}\Omega_n\Psi_n^{-1}$ برآوردی از کارایی مجانبیِ برآوردگرهای حداقل واگرایی توان چگالی $MDPDE$ ها $(\hat{\boldsymbol{\beta}}_n,\hat{\phi}_n)$ را ارائه می‌دهد. اگرچه این ،بستگی به‌اندازه نمونه $n$ و متغیرهای $x_i$ داده‌شده دارد ، اما برآورد

معقولی از کارایی مجانبی برای مقادیرِ بزرگِ $n$ نشان می‌دهد. علاوه بر این ، کواریانس مجانبی برآوردگرهای $\hat{\beta}_n$ و $\hat{\phi}_n$ به‌طورکلی صفر نیست و بنابراین این برآوردگرها برای همه مدل‌های خطی تعمیم‌یافته ($GLM$ها) مجانباً مستقل نیستند.

## ۳ اجرای روش $MDPDE$ در برآورد ضرایب مدل رگرسیون پواسون برای داده‌های شمارش‌پذیر

در تحقیقات و متون آماری قدیمی، هیچ روش توانمند خاصی برای برآورد پارامترهای رگرسیون پواسون وجود ندارد. بااین‌حال، در این بخش از روش توانمند برآوردگر حداقل واگرایی توان چگالی برای برآورد ضرایب مدل رگرسیون پواسون استفاده خواهیم کرد؛ بنابراین بر روی رگرسیون پواسون تمرکز می‌کنیم که در آن مقادیر مشاهده‌شده متغیرهای پاسخ مجموعه‌ای از اعداد صحیح غیر منفی هستند. مفیدترین ابزار رگرسیون برای داده‌های شمارش‌پذیر، مدل رگرسیون پواسون است که در آن ، با توجه به مقادیر متغیرهای توضیحی، متغیرهای پاسخ به‌طور مستقل از توزیع پواسون پیروی می‌کنند اما این توزیع بسته به مقادیر متناظر متغیر توضیحی ممکن است پارامترهای میانگین متفاوتی داشته باشد.

به‌طور دقیق‌تر، فرض کنید $(x_۱, y_۱), \ldots, (x_n, y_n)$ مشاهدات نمونه از مدل رگرسیون پواسون باشد؛ یعنی یک مدل خطی تعمیم‌یافته در نظر می‌گیریم که $Y_i$ ها متغیرهای پاسخ با توزیع پواسون و $X_i$ ها متغیرهای مستقل توضیحی هستند. همچنین فرض کنید که مقادیر متغیر توضیحی $x_i$ ثابت است. پس در مدل رگرسیون پواسون، متغیرهای پاسخِ شمارش‌پذیر $Y_i$ مستقل بوده و دارای توزیع پواسون با پارامتر $\mu_i = e^{(\boldsymbol{x}_i^T \boldsymbol{\beta})}$ و

$$\mu_i = E(Y_i|x_i) = e^{(\boldsymbol{x}_i^T \boldsymbol{\beta})}$$

هستند. به‌طوری‌که $\mu_i > ۰$ و میانگین و واریانس متناظر با هرکدام از $Y_i$ ها یکسان و برابر با مقدار $\mu_i$ است. پس

$$\mu_i = E(Y_i|x_i) = e^{(\boldsymbol{x}_i^T \boldsymbol{\beta})} \quad , \quad Var(Y_i|x_i) = e^{(\boldsymbol{x}_i^T \boldsymbol{\beta})}$$

بنابراین در این مدل خطی تعمیم‌یافته به ازای هر $i$ و $i = ۱, \ldots, n$ ، $Y_i \sim PO(\mu_i)$. یعنی $Y_i$ ها متغیرهای تصادفی مستقل و غیرهم‌توزیع و با تابع چگالی احتمال زیر می‌باشند:

$$f(y_i; \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \quad y_i = ۰, ۱, ۲, ۳, \ldots \qquad (۲۵)$$

این تابع چگالی را می‌توان به فرم خانواده توزیع‌های نمایی (و حالت خاص آن یعنی فرم کانونی) به‌صورت زیر نوشت:

$$f(y_i; \mu_i) = exp\left\{y_i \log(\mu_i) - \mu_i - \log(y_i!)\right\} \qquad (۲۶)$$

با نمادگذاری $\eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$ و $\mu_i = e^{(\boldsymbol{x}_i^T \boldsymbol{\beta})} = e^{\eta_i}$ فرم کانونی تابع چگالی به‌صورت زیر خواهد بود:

$$\begin{aligned} f(y_i; \eta_i) &= exp\left\{y_i \log(e^{\eta_i}) - e^{\eta_i} - \log(y_i!)\right\} \\ &= exp\left\{y_i \eta_i - e^{\eta_i} - \log(y_i!)\right\} \qquad (۲۷) \end{aligned}$$

ملاحظه می‌کنیم که رگرسیون پواسون در واقع یک مورد خاص از $GLM$ با پارامتر شکل معلومِ $\phi = ۱$ و $\theta_i = \eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$ ، $b(\theta_i) = e^{\theta_i} = e^{\eta_i}$ ، $c(y_i) = -\log(y_i!)$ و $a(\phi) = ۱$ است.

فرض کنید $g$ یک تابع یکنوای صعودی با دامنه اعداد حقیقی مثبت باشد. ازآنجایی‌که در اینجا میانگین $\mu_i = e^{(\boldsymbol{x}_i^T \boldsymbol{\beta})} = e^{\eta_i}$ است، تابع پیوند $g$ را تابع لگاریتم طبیعی " log " انتخاب می‌کنیم. لذا پارامترهای مدل رگرسیونی ما $\boldsymbol{\beta}$ از طریق رابطه زیر وارد مدل می‌شوند:

$$g(\mu_i) = \log(\mu_i) = \boldsymbol{x_i^T \beta} = \eta_i$$

می‌دانیم $x_i^T = (x_{i۱}, \ldots, x_{ip})$ بردار متغیرهای مستقل مربوط به $i$اُمین مشاهده و $\boldsymbol{\beta} = (\beta_۱, \ldots, \beta_p)^T$ بردار ضرایب رگرسیونی هستند. در بخش قبلی نشان دادیم که در حالتی که داده‌ها مستقل و غیرهم‌توزیع هستند و حالتی که مقدار پارامتر $\phi$ معلوم و ثابت باشد می‌توانیم برآورد حداقل واگرایی توان چگالی $\boldsymbol{\beta}$ را به‌وسیله مینیمم کردن عبارت زیر به دست آوریم:

$$\sum_{i=۱}^{n} x_i \left[ \int K_{۱i}(y; \boldsymbol{\beta}) f_i(y; \boldsymbol{\beta})^{۱+\alpha} dy - K_{۱i}(y_i; \boldsymbol{\beta}) f_i(y_i; \boldsymbol{\beta})^{\alpha} \right] = ۰ \qquad (۲۸)$$

یعنی هدف ما پیدا کردن آن مقدار از بردار پارامتری $\boldsymbol{\beta}$ است که به ازای آن فاصله بین چگالی‌های مدل پیشنهادشده و چگالی‌های واقعی برای تمام نقاط حداقل شود. در قضیه زیر نحوه محاسبه برآوردگر حداقل واگرایی توان چگالی مربوط به مدل رگرسیون خطی تعمیم‌یافته ارائه می‌شود:

**قضیهٔ ۱.۳.** *فرض کنیم یک نمونه تصادفی به‌صورت* $(x_i, y_i)$ *و* $i = ۱, \ldots, n$ *داریم، اگر مدل رگرسیون خطی تعمیم‌یافته پس از اعمال تابع*

*پیوند لگاریتمی را به‌صورت*

$$g(\mu_i) = \log(\mu_i) = \boldsymbol{x_i}^T\boldsymbol{\beta}, \quad i = ۱, \ldots, n$$

*در نظر بگیریم که متغیرهای پاسخ شمارش‌پذیر $y_i$ دارای توزیع پواسون با پارامتر $\mu_i = e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})}$ هستند و $x_i^T$ بردار متغیرهای مستقل مربوط به $i$اُمین مشاهده و $\beta$ بردار ضرایب رگرسیونی باشند. آنگاه با استفاده از نمادگذاری بالا و شکل توزیع پواسون، برآوردگر حداقل واگرایی توان چگالی مربوط به ضرایب رگرسیونی $\beta_j$ ، $j = ۱, \ldots, p$ برای $\alpha \geq ۰$ از حل کردن معادله زیر به دست می‌آید:*

$$\sum_{i=۱}^{n}\Big[\ \gamma_{۱i}(\boldsymbol{\beta}) - \ (y_i - \ e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})})f_i(y;\boldsymbol{\beta})^\alpha\Big]\, x_i = ۰\ ,\ j = ۱, ..., p \tag{۲۹}$$

*به‌طوری‌که $f_i(y;\beta)$ تابع چگالی احتمال توزیع پواسون با میانگین $\mu_i = e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})}$ است و*

$$\gamma_{۱i} = \gamma_{۱i}^{۱+\alpha}(\boldsymbol{\beta}) = \int K_{۱i}(y;\boldsymbol{\beta})f_i(y;\boldsymbol{\beta})^{۱+\alpha}dy$$

اثبات قضیه بالا در پیوست ۱ آمده است.

**نتیجه ۲.۳.** *به‌طور خاص، برای $\alpha = ۰$ ، معادله برآوردی بالا به معادله برآوردی ماکسیمم درستنمایی $MLE$ ارائه‌شده توسط رابطه زیر ساده می‌شود.*

$$\sum_{i=۱}^{n}(y_i - \ e^{(\boldsymbol{x_i}^T\boldsymbol{\beta})})x_i \ = \ ۰ \tag{۳۰}$$

بااین‌حال، برای $\alpha > ۰$ ، هیچ شکل ساده‌شده‌ای برای $\gamma_{۱i}$ و $\gamma_{۱۱i}$ وجود ندارد، بنابراین ما نیاز داریم که این کمیت‌ها را به‌صورت عددی محاسبه و سپس معادله برآوردی ۴۴ را برحسب مؤلفه‌های بردار ضرایب رگرسیونی $\beta$ به‌صورت عددی حل کنیم.

## ۱.۳ ویژگی‌های مجانبی برآوردگر $MDPDE$ مربوط به پارامترهای مدل رگرسیون پواسون

ویژگی‌های مجانبی برآوردگر $MDPDE$ برای پارامترهای $\beta$ تحت مدل رگرسیون پواسون مستقیماً از قضیه ۱۰.۲ پیروی می‌کند. برای بررسی خواص مجانبی برآوردگرها ابتدا قضیه زیر را بیان می‌کنیم.

**قضیهٔ ۳.۳.** *فرض کنیم که توزیع واقعی مولد داده‌ها زیرمجموعه خانواده مدل توزیع‌ها باشد به‌طوری‌که $g_i(y) = f_i(y;\boldsymbol{\theta})$ و یک نمونه تصادفی مستقل و غیرهم‌توزیع به‌صورت $(x_i, y_i)$ و $i = ۱, \ldots, n$ داریم، همچنین مدل رگرسیون خطی تعمیم‌یافته پس از اعمال تابع پیوند لگاریتمی را به‌صورت*

$$g(\mu_i) = \log(\mu_i) = \boldsymbol{x_i}^T\boldsymbol{\beta}, \quad i = ۱, \ldots, n$$

*در نظر می‌گیریم که متغیرهای پاسخ $Y_i \sim PO(e^{(x_i^T\boldsymbol{\beta})})$ ، دارای توزیع پواسون با پارامتر $\mu_i = e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})}$ هستند و $x_i^T$ بردار متغیرهای مستقل مربوط به $i$اُمین مشاهده و $\beta$ بردار ضرایب رگرسیونی باشند. آنگاه مقادیر $N_i^{۱+\alpha}$، $M_i^{۱+\alpha}$ ، $\Psi_n$ و $\Omega_n$ به‌صورت زیر به دست می‌آید:*

$$N_i^{۱+\alpha}(\boldsymbol{\beta}) = \sum_{y=۰}^{\infty}(y - \ e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})})x_i f_i(y;\boldsymbol{\beta})^{۱+\alpha} \ = \gamma_{۱i}x_i \tag{۳۱}$$

$$\begin{aligned} M_i^{۱+\alpha}(\boldsymbol{\beta}) &= \sum_{y=۰}^{\infty}(y - \ e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})})^۲(x_i x_i^T)f_i(y;\boldsymbol{\beta})^{۱+\alpha} \\ &= \gamma_{۱۱i}(x_i x_i^T) \end{aligned} \tag{۳۲}$$

$$\Psi_n(\boldsymbol{\beta}) = \frac{۱}{n}\left(X^T\Gamma_{۱۱}^{(\alpha)}X\right) \tag{۳۳}$$

$$\Omega_n(\boldsymbol{\beta}) = \frac{۱}{n}\left(X^T[\Gamma_{۱۱}^{(۲\alpha)} - (\Gamma_{۱}^{(\alpha)})^۲]X\right) \tag{۳۴}$$

*به‌طوری‌که $f_i(y;\beta)$ تابع چگالی احتمال توزیع پواسون با میانگین $\mu_i = e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})}$ است.*

*اثبات.* با توجه به گسسته بودن متغیر تصادفی پواسون می‌توان در تعاریف و نمادگذاری‌های بخش قبلی، به‌جای علامت $\int$ از علامت $\sum$ و جمع روی تمام مقادیر متغیر تصادفی پواسون استفاده کرد. سپس می‌توان با استفاده از فرم تابع چگالی پواسون و تعریف $u_i(y;\beta)$ برحسب $K_{۱i}(y;\beta)$ مطابق با رابطه ۱۴ مقدار $K_{۱i}(y;\beta)$ و در نهایت مقدار $\gamma_{۱i}$ از رابطه ۴۰ مقادیر $N_i^{۱+\alpha}$ و $M_i^{۱+\alpha}$ را به دست آورد. با جایگذاری $\boldsymbol{M}_i^{۱+\alpha}$ ها و $\boldsymbol{N}_i^{۱+\alpha}$ ها با روابط و تعاریف متناظر آن‌ها و انجام محاسبات ساده جبری مقادیر $\Psi_n$ و $\Omega_n$ به دست می‌آیند. □

**تذکر ۴.۳.** *نکته قابل‌توجه این است که شرایط روی $x_i$ به‌گونه‌ای است که مقادیر باقیمانده در نمونه‌های بزرگ کران‌دار هستند. همچنین طیفی از باقیمانده‌های ماتریس مجموع فراهم‌شده‌ی $(\mathbf{X}^T\mathbf{X})$ ، دور از نقطه صفر کران‌دار هستند.*

## ۲.۳ توزیع مجانبی برآوردگر $MDPDE$ مربوط به پارامترهای مدل رگرسیون پواسون

**قضیهٔ ۵.۳.** *تحت تنظیمات مربوط به مدل رگرسیون پواسون که در ابتدای این بخش مطرح شد و همچنین تحت فرض‌های ۱ تا ۷ گوش و باسو (۲۰۱۳) نتایج زیر اتفاق می‌افتد:*

*۱) بازه سازگاری به‌عنوان $\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_n^{(\alpha)}$ از ریشه‌های معادله برآوردی حداقل $DPD$ در رابطه ۲۹ برای پارامتر همسان‌ساز $\alpha$ وجود دارد.*

*۲)* $$(\mathbf{X}^T[\Gamma_{11}^{(2\alpha)}(\boldsymbol{\beta}^g) - (\Gamma_1^{(\alpha)})^2(\boldsymbol{\beta}^g)]\mathbf{X})^{\frac{-1}{2}} (\mathbf{X}^T\Gamma_{11}^{(\alpha)}(\boldsymbol{\beta}^g)\mathbf{X})(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^g)$$

*دارای توزیع مجانبی نرمال $p$ بعدی با بردار میانگین صفر و ماتریس واریانس–کوواریانس $\boldsymbol{I}_p$ می‌باشد.*

*واریانس مجانبی برآوردگر $\beta^g$ را به‌اختصار با $AV$ [19] نشان می‌دهیم و به‌صورت زیر اندازه‌گیری می‌شود*

$$AV_\alpha(\boldsymbol{\beta}^g) = (\mathbf{X}^T\Gamma_{11}^{(\alpha)}(\boldsymbol{\beta}^g)\mathbf{X})^{-1}(\mathbf{X}^T[\Gamma_{11}^{(2\alpha)}(\boldsymbol{\beta}^g) - (\Gamma_1^{(\alpha)})^2(\boldsymbol{\beta}^g)]\mathbf{X})(\mathbf{X}^T\Gamma_{11}^{(\alpha)}(\boldsymbol{\beta}^g)\mathbf{X})^{-1} \quad (۳۵)$$

کارایی مجانبی برآوردگرهای مختلف $MDPDE$ برای $\boldsymbol{\beta}$ ،یعنی $\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_n^{(\alpha)}$ ،را می‌توان بر اساس واریانس مجانبی فوق اندازه‌گیری کرد. به‌طوری‌که می‌توان واریانس مجانبی این برآوردگرها را به‌طور مداوم با جایگزینی $\boldsymbol{\beta}^g$ با $\hat{\boldsymbol{\beta}}_n$ برآورد کرد، یعنی:

$$\widehat{AV}_\alpha = AV_\alpha(\hat{\boldsymbol{\beta}}_n)$$

نکته قابل‌توجه دیگر این است که برآوردگر حداقل واگرایی توان چگالی در حالت $\alpha = 0$ همان برآوردگر ماکزیمم درستنمایی است؛ بنابراین اگر در عبارات مربوط به واریانس مجانبی مقدار $\alpha$ را برابر صفر قرار دهیم، آنگاه واریانس مجانبی برآوردگر $MLE$ یعنی $AV_0$ به دست می‌آید.

## ۳.۳ کارایی نسبی مجانبی برآوردگر $MDPDE$ مربوط به پارامترهای مدل رگرسیون پواسون

کارایی نسبی مجانبی برآوردگر را به‌اختصار با $ARE$ [20] نشان می‌دهیم که در اینجا $ARE$ کارایی مجانبی برآوردگر حداقل واگرایی توان چگالی نسبت به کارایی برآوردگر $MLE$ به دست می‌آید (می‌دانیم که برآوردگر ماکزیمم درستنمایی دارای حداکثر میزان کارایی ($e_n(T_n) = 1$) است)؛ بنابراین، برآوردی از کارایی نسبی $MDPDE$های مختلف از مؤلفه $i$اُم بردار پارامتری $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ با توجه به کارایی برآوردگر $MLE$ آن (برآوردگر $OLS$ ) به‌صورت زیر ارائه می‌شود.

$$\widehat{RE}_{i,\alpha} = \frac{\widehat{AV}_0^i}{\widehat{AV}_\alpha^i} \times 100$$

که $\widehat{AV}^i$ در واقع $i$–مین مؤلفه قطریِ ماتریس $\widehat{AV}$ است. واضح است که برآورد بالا از کارایی نسبی، به حجم نمونه $n$ و انتخاب متغیرهای توضیحی داده‌شده $x_i$ بستگی دارد؛ اما می‌توان نشان داد که اگر $x_i$ ها مناسب انتخاب شود آنگاه سازگاری برآوردگر $\hat{\boldsymbol{\beta}}_n$ نشان‌دهنده این است که اندازه‌گیری بالا برآوردگر سازگاری از کارایی نسبی مجانبی به ما می‌دهد. به‌عنوان‌مثال، $\mathbf{X}^T\mathbf{X}$ باید محدود باشد.

**تذکر ۶.۳.** *برآوردگر حداقل واگرایی توان چگالی به ازای $\alpha = 0$ همان برآوردگر درستنمایی ماکزیمم است که دارای بالاترین میزان کارایی یعنی مقدار ۱ است. با افزایش میزان $\alpha$ از کارایی برآوردگر حداقل واگرایی توان چگالی کاسته می‌شود و هم‌زمان میزان توانمندی این برآوردگر افزایش می‌یابد. با انجام شبیه‌سازی مشخص می‌شود که از دست دادن کارایی برای $MDPDE$ با $\alpha$ مثبت کوچک کاملاً ناچیز است؛ بنابراین میزان کاهش $ARE$ به ازای مقادیر کوچک $\alpha$، بسیار کم است. حتی برای $\alpha$ مثبت بزرگ نزدیک به ۰٫۵، اگر $x_i$ ها نسبتاً کوچک باشد، می‌توانیم کارایی بسیار بالایی داشته باشیم.*

# ۴ مدل‌سازی داده‌های واقعی به روش $MDPDE$ در مدل رگرسیون پواسون

در این بخش، عملکرد $MDPDE$های پیشنهادی را در مدل رگرسیون پواسون و با اعمال آن بر روی یک مجموعه داده واقعی بررسی خواهیم کرد و نشان خواهیم داد که برآوردهای حاصل از این روش در مقابل مشاهدات پرت و دارای نفوذ، توانمند بوده و کارایی قابل قبولی دارند.

[19] Asymptotic Variance

[20] Asymptotic Relative Efficiency

## ۱.۴ مدل‌سازی داده‌های صرع به روش $MDPDE$ در مدل رگرسیون پواسون

**مثال ۱.۴.** *در این مثال یک مجموعه داده شامل ۵۹ بیمار صرعی از تال و ویل (۱۹۹۰) [۱۶] (جدول ۲، بخش ۵) را در نظر می‌گیریم. داده‌ها از یک کار آزمایی بالینی انجام‌شده توسط لیپک و همکاران (۱۹۸۵) به دست آمد. به‌طوری‌که در آن کار آزمایی بیماران مبتلا به صرع با داروی ضد صرع «پروگابید» یا با «دارونما (تسکین‌دهنده)» با انتساب تصادفی تحت درمان قرار گرفتند. در واقع تقریباً نیمی از بیماران (۲۸ بیمار) به‌صورت تصادفی انتخاب شدند و با استفاده از «دارونما (تسکین‌دهنده)» تحت درمان قرار گرفتند و نیمی دیگر از بیماران (۳۱ بیمار) با استفاده از داروی ضد صرع «پروگابید» تحت درمان قرا گرفتند.*

در این مثال تعداد کل حملات صرع ذکرشده در طول چهار دوره برای هر بیمار را باهم جمع می‌کنیم و به‌عنوان متغیر پاسخ $Y_i$ در نظر می‌گیریم که با توجه به نحوه پراکندگی مشاهدات حاصل از این متغیر، می‌توان گفت که متغیر پاسخ $Y_i$ از توزیع پواسون پیروی می‌کند؛ بنابراین مجموعه داده‌های صرع را با مجموعه‌ای از متغیرهای توضیحی مناسب از طریق مدل رگرسیون پواسون مدل‌سازی می‌کنیم (حسینیان ۲۰۰۹ [۱۱]). متغیرهای توضیحی در نظر گرفته‌شده در این رابطه عبارت‌اند از: پایه [۲۱] ( نرخ تشنج پایه هشت‌هفته‌ای قبل از تصادفی سازی در مضرب ۴)، سن [۲۲] ( سن بیمار در مضربی از ۱۰ سال) و درمان کمکی [۲۳] ( یک شاخص باینری برای گروه کنترل-درمان). توجه کنید که تعامل بین درمان و نرخ تشنج پایه در این مثال مهم است، زیرا بسته به تعداد پایه، این تعامل نرخ تشنج بالاتر یا پایین‌تری را برای گروه درمان در مقایسه با گروه دارونما نشان می‌دهد. در واقع، این دارو تنها در صورتی صرع را کاهش می‌دهد که تعداد پایه با توجه به آستانه بحرانی به‌اندازه کافی زیاد شود. این داده‌ها توسط حسینیان (۲۰۰۹) [۱۱] تجزیه‌وتحلیل شدند که برآوردگر ماکسیمم درستنمایی [۲۴] را با روش‌های توانمند پیشنهادشده توسط خود او(برآوردگر ماکسیمم درستنمایی وزنی [۲۵] ) در همان مقاله و روش‌های کانتونی و رانچتی (۲۰۰۱) [۶] مقایسه کرد. با توجه به نتایج مربوط به برآورد پارامترهای مدل رگرسیون پواسون و مقادیر $p$ آن‌ها برای مجموعه داده‌های صرع که توسط حسینیان و کانتونی ارائه‌شده و در جدول ۶-۴ آمده است می‌توان گفت که با استفاده از روش‌های $WMLE$ (حسینیان) و $MLE$ پارامترهای $\beta_1$ ، $\beta_2$ و $\beta_3$ در مدل ۳۷ می‌مانند اما چالش و تفاوت در روش‌های توانمند (مانند حسینیان) و $MLE$ بر روی این است که آیا پارامتر $\beta_4$ نیز در مدل می‌ماند یا خیر؟

حسینیان بر این باور است که به دلیل مشاهدات پرت، اثر متقابل بین درمان و نرخ تشنج پایه بر اساس برآوردگر ماکسیمم درستنمایی ناچیز است؛ یعنی چون اندازه $p$-مقدار برای $\beta_4$ (ضریب اثر متقابل $Trt : Base$ ) بر اساس برآوردگر $MLE$ نسبتاً بزرگ است پس فرض $H_0$ در آزمون فرض ۳۶ رد نمی‌شود و $\beta_4$ در مدل نمی‌ماند. درحالی‌که بر اساس برآوردگرهای پیشنهادی توانمندِ حسینیان ($WMLE$) اندازه $p$-مقدار برای $\beta_4$ خیلی کوچک است پس فرض $H_0$ در آزمون فرض ۳۶ رد می‌شود و لذا ضریب اثر متقابل $Trt : Base$ یعنی پارامتر $\beta_4$ در مدل می‌ماند؛ بنابراین حسینیان این تعامل را معنی‌دار نشان می‌دهد و با استفاده از روش پیشنهادی خودشان متغیر متقابل را در مدل نگه می‌دارد.

پس نتیجه می‌گیریم چنانچه مدل توانمندی به داده‌ها برازش داده شود، انتظار داریم که پارامتر $\beta_4$ معنی‌دار شود. لذا در این بخش از مقاله، برآوردگرهای حداقل واگرایی توان چگالی توانمند را برای مجموعه داده‌های صرع اعمال می‌کنیم و سعی می‌کنیم نشان دهیم که آیا برآوردگرهای پیشنهادی ما نیز به‌اندازه کافی توانمند هستند که با برآوردگر ماکسیمم درستنمایی برای اثر متقابل $Trt : Base$ متمایز شوند؟ آیا روش $MDPDE$ بهتر از روش $MLE$ عمل می‌کند؟ آیا مانند روش حسینیان($WMLE$) اثر متقابل $Trt : Base$ را در مدل نگه می‌دارد؟

برای پیدا کردن پاسخ، سؤال موردعلاقه فوق را می‌توان به‌عنوان آزمونی برای فرضیه‌های زیر فرمول‌بندی کرد:

$$\begin{cases} H_0 : \beta_4 = 0 \\ H_1 : \beta_4 \neq 0 \end{cases} \qquad (۳۶)$$

برای داده‌ها مدل رگرسیون پواسون را در نظر می‌گیریم به‌طوری‌که لگاریتم نرخ تشنج پایه‌ای ($Base$) ، لگاریتم سن ($Age$) و شاخص باینری ($Trt$) به‌عنوان متغیر مستقل و لگاریتم تعداد کل تشنج در هر چهار معاینه بالینی متوالی پس از تصادفی سازی ($Y_i$) به‌عنوان متغیر

[21] Base
[22] Age
[23] Trt
[24] MLE
[25] WMLE

وابسته در نظر گرفته‌شده است.

$$\text{متغیرهای مستقل} = \begin{cases} X_۱ = Trt \\ X_۲ = Base \\ X_۳ = Age \\ X_۴ = Trt \times Base \end{cases}$$

و

$$\text{متغیر وابسته} = Y_i$$

تابع پیوند $g$ را تابع لگاریتم طبیعی "log" انتخاب می‌کنیم. لذا پارامترهای مدل رگرسیونی پواسون $\boldsymbol{\beta}$ از طریق رابطه زیر وارد مدل می‌شوند:

$$g(\mu_i) = \log(\mu_i) = \boldsymbol{x_i^T \beta} = \eta_i$$

بنابراین مدل رگرسیون پواسون برای داده‌های صرع به‌صورت زیر خواهد بود:

$$\log(\hat{\mu}) = \log(E(\hat{Y})) = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} = \hat{\beta}_۰ + \hat{\beta}_۱ X_۱ + \hat{\beta}_۲ X_۲ + \hat{\beta}_۳ X_۳ + \hat{\beta}_۴ X_۴ \qquad (۳۷)$$

نتایج مربوط به برآورد پارامترهای مدل رگرسیون پواسون ۳۷ ، خطاهای استاندارد و مقادیر $p$ آن‌ها برای مجموعه داده‌های صرع با استفاده از روش برآورد ماکسیمم درستنمایی و روش برآورد ماکسیمم درستنمایی وزنی در جدول ۱ و بر اساس برآوردگر حداقل واگرایی توان چگالی با $\alpha$های متفاوت در جدول ۲ ارائه‌شده است. در جدول ۲ ملاحظه می‌کنیم بر اساس برآوردگر توانمندِ حداقل واگرایی توان چگالی برای $\alpha = ۰/۳$ ، $\alpha = ۰/۵$ و $\alpha = ۰/۷$ اندازه $p$-مقدار برای $\beta_۴$ خیلی کوچک است پس فرض $H_۰$ در آزمون فرض ۳۶ رد می‌شود و لذا ضریب اثر متقابل $Trt : Base$ یعنی پارامتر $\beta_۴$ در مدل می‌ماند؛ بنابراین روش $MDPDE$ این تعامل را معنی‌دار نشان می‌دهد و استفاده از روش پیشنهادی توانمند ما متغیر متقابل را در مدل نگه می‌دارد.

نتیجه: واضح است که برآوردگرهای $MDPDE$ مربوط به $\alpha \geq ۰/۳$ با برآوردگر ماکسیمم درستنمایی کاملاً متفاوت هستند. با توجه به نتایج به‌دست‌آمده ملاحظه می‌گردد روش برآورد حداقل واگرایی توان چگالی برای $\alpha \geq ۰/۳$ یک روش توانمند می‌باشد زیرا برای $\alpha \geq ۰/۳$ ضریب $\beta_۴$ معنادار شده است. بنابراین برای برآوردگرهای $MDPDE$ اثر متقابل تحت مدل رگرسیون پواسون نیز قابل‌توجه است. در واقع، این برآوردگرها کاملاً شبیه به برآوردگرهای توانمند در نظر گرفته‌شده در حسینیان (۲۰۰۹) هستند، اما همان‌طور که قبلاً توضیح دادیم، دارای خواص مجانبی برتری هستند زیرا همان‌طور که در جداول ۱و ۲ ملاحظه می‌کنید با توجه به انحراف استاندارد ضرایب برآورد شده از روش حسینیان و روش $MDPDE$، با $\alpha = ۰/۱$ روش $MDPDE$ روشی کاراتر می‌باشد.

# پیوست

پیوست ۱: به ازای هر نقطه از مشاهدات ،چگالی مدل پیشنهادی به‌صورت $f_i \equiv PO(\mu_i)$ است و داریم:

$$f_i \sim PO(\mu_i) \Longrightarrow f(y_i;\mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \quad y_i = ۰, ۱, ۲, ۳, \ldots \qquad (۳۸)$$

در بخش ۲ برای $i = ۱, \ldots, n$ و $j, k = ۱, ۲$ مقدار $\gamma_{ji}$ را به‌صورت عبارت زیر تعریف کردیم:

$$\gamma_{ji} = \gamma_{ji}^{۱+\alpha}(\boldsymbol{\beta}, \phi) = \int K_{ji}(y;(\boldsymbol{\beta},\phi)) f_i(y;(\boldsymbol{\beta},\phi))^{۱+\alpha} dy \qquad (۳۹)$$

چون پارامتر $\phi$ معلوم و ثابت است آن را در رابطه کنار می‌گذاریم. پس برای $j = ۱$ مقدار $\gamma_{۱i}$ به‌صورت عبارت زیر خواهد بود:

$$\gamma_{۱i} = \gamma_{۱i}^{۱+\alpha}(\boldsymbol{\beta}) = \int K_{۱i}(y;\boldsymbol{\beta}) f_i(y;\boldsymbol{\beta})^{۱+\alpha} dy \qquad (۴۰)$$

اکنون جهت اثبات قضیه از رابطه ۲۸ به‌صورت زیر استفاده می‌کنیم:

$$\sum_{i=۱}^{n} x_i \left[ \int K_{۱i}(y;\boldsymbol{\beta}) f_i(y;\boldsymbol{\beta})^{۱+\alpha} dy - K_{۱i}(y_i;\boldsymbol{\beta}) f_i(y_i;\boldsymbol{\beta})^{\alpha} \right] = ۰$$

ملاحظه می‌کنیم که قسمت اول عبارت فوق برابر با رابطه ۴۰ است، بنابراین با جایگذاری مقدار $\gamma_{۱i}$ در رابطه فوق داریم:

$$\sum_{i=۱}^{n} x_i \left[ \gamma_{۱i}(\boldsymbol{\beta}) - K_{۱i}(y_i;\boldsymbol{\beta}) f_i(y_i;\boldsymbol{\beta})^{\alpha} \right] = ۰ \qquad (۴۱)$$

با توجه به رابطه ۱۴ مقدار $K_{۱i}$ به‌صورت زیر تعریف می‌شود:

$$K_{۱i}(y_i;\boldsymbol{\beta}) = \frac{(y_i - \mu_i)}{Var(Y_i) g'(\mu_i)} \qquad (۴۲)$$

از طرفی

$$g(\mu_i) = \log(\mu_i) \Longrightarrow g'(\mu_i) = \frac{۱}{\mu_i}, \quad Var(Y_i) = \mu_i$$

حال مقادیر میانگین و واریانس را در رابطه ۴۲ جایگذاری می‌کنیم

$$K_{۱i}(y_i;\boldsymbol{\beta}) = \frac{(y_i - \mu_i)}{\mu_i . \frac{۱}{\mu_i}} = (y_i - \mu_i) = (y_i - e^{(\boldsymbol{x}_i^T \boldsymbol{\beta})}) \qquad (۴۳)$$

| | Int. | Age | Base | Trt | Base:Trt |
|---|---|---|---|---|---|
| MLE | ۱٫۹۷ | ۰٫۲۴ | ۰٫۰۸۵ | $-$۰٫۲۶ | ۰٫۰۰۸ |
| st.err | ۰٫۱۳ | ۰٫۰۴ | ۰٫۰۰۴ | ۰٫۰۷۷ | ۰٫۰۰۴ |
| p-value | ۲$e-$۱۶ | ۳٫۷$e-$۹ | ۲$e-$۱۶ | ۰٫۰۰۰۹ | ۰٫۰۸۸ |
| WMLE | ۲٫۱۳ | ۰٫۰۴۴ | ۰٫۱۲۸ | $-$۰٫۴۷ | ۰٫۰۵۴ |
| st.err | ۰٫۱۹۸ | ۰٫۰۵۷ | ۰٫۰۱۴ | ۰٫۱۶ | ۰٫۰۲۱ |
| p-value | ۱٫۱$e-$۲۷ | ۴٫۴$e-$۰۱ | ۹٫۲$e-$۲۱ | ۲٫۶$e-$۰۳ | ۱$e-$۰۲ |

جدول ۱: برآورد پارامترهای مدل رگرسیون پواسون برای مجموعه داده‌های صرع از روش ماکسیمم درستنمایی ($MLE$) و روش ماکسیمم درستنمایی وزنی ($WMLE$) با خطاهای استاندارد و مقادیر p آن‌ها.

در نهایت با جایگذاری $K_{1i}(y_i;\beta)$ در رابطه ۴۱ قضیه اثبات می‌شود:

$$\sum_{i=1}^{n}\left[\gamma_{1i}(\boldsymbol{\beta})-\ (y_i-\ e^{(\boldsymbol{x}_i^T\boldsymbol{\beta})})f_i(y;\boldsymbol{\beta})^{\alpha}\right]x_i\ =\ 0 \qquad (۴۴)$$

ریشه‌های معادله ۴۴ همان مقادیر برآوردهایی است که به دنبال به دست آوردن آن‌ها بودیم. بسته به تعداد متغیرهای مستقل به‌کاررفته در مدل رگرسیون ($p$) با حل کردن این معادله و به دست آوردن ریشه‌های آن‌ها برآورد بردار پارامتری $\beta$ به دست می‌آید.

# مراجع


[1] Abonazel, M. R., Saber, O. M. (2020). A comparative study of robust estimators for Poisson regression model with outliers. *Journal of Statistical Applications and Probability*, **9**, 279-286 .

[2] Abonazel, M. R., Dawoud, I. (2022). Developing robust ridge estimators for Poisson regression model. *Concurrency and Computation: Practice and Experience*, **34(15)**, e6979 .

[3] Amin, M., Akram, M. N., Kibria, B. G. (2021). A new adjusted Liu estimator for the Poisson regression model. *Concurrency and computation: Practice and experience*, **33(20)**, e6340 .

[4] Basu,A., Harris,I.R., Hjort, N.L. and Jones M.C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*. **85**, 3, 549–559.

[5] Beran,R. (1977). Minimum Hellinger Distance Estimates for Parametric Models. *The annals of Statistics*, **5**, 3, 445-463.

[6] Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalised linear models. *Journal of the American Statistical Association*. **96**, 455, 1022–1030.

[7] Chen, W., Qian, L., Shi, J., Franklin, M. (2018). Comparing performance between log-binomial and robust Poisson regression models for estimating risk ratios under model misspecification. *BMC medical research methodology*, **18(1)**, 1-12 .

[8] Durio, A. and Isaia, E.D. (2011). The Minimum Density Power Divergence Approach in Building Robust Regression Models. *INFORMATICA*. **22**, 1, 43–56.

| | $\alpha =$ ۰ | $\alpha =$ ۰/۱ | $\alpha =$ ۰/۳ | $\alpha =$ ۰/۵ | $\alpha =$ ۰/۷ | $\alpha =$ ۱ |
|---|---|---|---|---|---|---|
| Intercept | | | | | | |
| Estimate | ۱/۹۶۸۰ | ۱/۹۵۷۸ | ۱/۶۳۸۹ | ۱/۷۳۵۳ | ۱/۶۷۷۱ | ۱/۸۱۶۹ |
| SE($\times$100) | ۱۳/۵۹۲۹ | ۱۵/۲۵۰۹ | ۱۲/۶۸۶۹ | ۱۳/۷۰۸۱ | ۱۴/۹۱۸۵ | ۱۷/۰۰۴۳ |
| p-value | $<$ ۲$e -$ ۱۶ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۰۰ |
| Trt | | | | | | |
| Estimate | −۰/۲۵۵۳ | −۰/۲۶۵۵ | −۰/۲۲۸۶ | −۰/۳۰۸۹ | −۰/۲۴۸۱ | −۰/۸۸۹۴ |
| SE($\times$100) | ۷/۶۵۲۵ | ۸/۶۸۱۲ | ۷/۹۱۳۹ | ۸/۴۵۶۶ | ۹/۱۱۱۱ | ۱۰/۱۷۸۷ |
| p-value | ۰/۰۰۰۸ | ۰/۰۰۰۶ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۰۳ | ۰/۰۰۲۷ |
| Base | | | | | | |
| Estimate | ۰/۰۸۵۴ | ۰/۰۵۴۷ | ۰/۰۸۵۰ | ۰/۱۵۲۴ | ۰/۰۸۵۶ | ۰/۱۵۴۹ |
| SE($\times$100) | ۰/۳۶۶۶ | ۰/۴۱۰۱ | ۰/۲۷۷۸ | ۰/۳۰۵۵ | ۰/۳۳۵۹ | ۰/۳۹۵۹ |
| p-value | $<$ ۲$e -$ ۱۶ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۰۰ |
| Age | | | | | | |
| Estimate | ۰/۲۴۳۵ | ۰/۲۰۵۶ | ۰/۲۴۲۲ | ۰/۱۳۸۶ | ۰/۲۳۶۴ | ۰/۱۰۴۴ |
| SE($\times$100) | ۴/۱۲۹۷ | ۴/۷۲۴۲ | ۳/۹۳۷۴ | ۴/۲۴۱۶ | ۴/۶۱۳۸ | ۵/۲۱۱۹ |
| p-value | ۳/۷۲$e -$ ۰۹ | ۰/۰۱۷۷ | ۰/۳۰۴۵ | ۰/۳۹۷۲ | ۰/۶۰۱۷ | ۰/۲۸۷۸ |
| Trt$\times$Base | | | | | | |
| Estimate | ۰/۰۰۷۵ | −۰/۱۰۷۶ | −۰/۰۶۷۹ | −۰/۰۶۸۹ | −۰/۰۶۸۵ | ۰/۰۶۴۰ |
| SE($\times$100) | ۰/۴۴۰۹ | ۰/۴۸۹۳ | ۰/۳۲۳۰ | ۰/۳۵۳۷ | ۰/۳۸۸۸ | ۰/۴۶۰۰ |
| p-value | ۰/۰۸۷۵ | ۰/۰۳۲۳ | ۰/۰۰۰۰ | ۰/۰۰۰۰ | ۰/۰۰۱۳ | ۰/۰۳۷۳ |

جدول ۲: برآورد پارامترهای مدل رگرسیون پواسون ، خطاهای استاندارد مجانبی و مقادیر $p$ آن‌ها برای مجموعه داده‌های صرع با استفاده از برآوردگر حداقل واگرایی توان چگالی با $\alpha$ های متفاوت.

[9] Ghosh, A. and Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electronic Journal of Statistics.* **7**, 2420–2456.

[10] Hall, D. B., Shen, J. (2010). Robust estimation for zero□inflated Poisson regression. *Scandinavian Journal of Statistics*, **37(2)**, 237-252 .

[11] Hosseinian, S. (2009). Robust inference for generalized linear models: binary and poisson regression. PHD thesis, Ecole Polytechnique Federal de Lausanne.

[12] Huber, P. J. (1983). Minimax Aspects of Bounded-Influence Regression (With Discussion). *Journal of the American Statistical Association*. **78**, 66-80.

[13] Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Annals of Statistics*. **1**, 799-821.

[14] Mallows, C. L. (1975). *on Some Topics in Robustness, Unpubhished Memorandom*. Bell Telephone Laboratories, Mary Hill, N. J.

[15] Simpson, D. G. (1987). Minimum Hellinger Distance Estimation for the Analysis of Count Data, *Journal of the American Statistical Association*, **82**, 802-807.

[16] Thall, P.F. and Vail S.C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*. **46**, 3, 657–671 .

[17] Tsou, T. S. (2006). Robust poisson regression. *Journal of Statistical Planning and Inference*, **136(9)**, 3173-3186 .

# Robust estimations in Poisson regression model

**Amirreza Mahmoodi[1] and Roohollah Roozegar[2]**

Abstract:

Many regression estimation techniques are strongly affected by outlier data and many errors occur in their estimation. In the recent years, robust methods have been developed to solve this issue. The minimum density power divergence estimator is an estimation method based on the minimum distance between two density functions, which provides a robust estimate in situations where the data contain a number of outliers. In this research, we present the robust estimation method of minimum density power divergence to estimate the parameters of the Poisson regression model, which can produce robust estimators with the least loss in efficiency. Also, we will investigate the performance of the proposed estimators by providing a real example.

**Keywords:** Generalized linear model, Minimum density power divergence estimator, Outlier data, Poisson Regression, Robustness, Tuning parameter .

[1] **Master's degree graduate, Collage of science, Yasouj university, Yasouj, Iran.**
[2] **Department of mathematics, Collage of science, Yasouj university, Yasouj, Iran**