دانشگاه علامه طباطبائی
دانشکده‌ی اقتصاد
گروه آمار، ریاضی و کامپیوتر

پایان‌نامه برای دریافت درجه‌ی کارشناسی ارشد **آمار ریاضی**

عنوان

تعیین براورد پارامترهای معادله‌های براوردساز تعمیم‌یافته درحضور داده‌های گم‌شده

پژوهشگر

مهدی عندلیب خواه

استاد راهنما

دکتر فرزاد اسکندری

استاد مشاور

دکتر حمیدرضا نواب پور

تیر ماه ۱۳۹۱


بسم الله الرحمن الرحيم

تأیید پایان‌نامه‌ی کارشناسی ارشدتوسط دانشجو

عنوان پایان‌نامه: تعیین براورد پارامترهای معادله‌های براوردساز تعمیم‌یافته درحضور داده‌های گم‌شده

نام دانشجو: مهدی عندلیب خواه
شماره‌ی دانشجویی: ۸۸۱۲۵۱۲۸۲۰۶
استاد راهنما: دکتر فرزاد اسکندری

این‌جانب **مهدی عندلیب خواه** دانشجوی کارشناسی ارشد رشته‌ی **آمار ریاضی** دانشکده‌ی اقتصاد دانشگاه علامه طباطبائی گواهی می‌نمایم پژوهش‌های ارائه شده در پایان‌نامه با عنوان مذکور توسط شخص این‌جانب انجام شده است و درستی مطالب نگارش یافته مورد تأیید می‌باشد. همچنین گواهی می‌نمایم مطالب مندرج در پایان‌نامه تاکنون برای دریافت هیچ نوع مدرک یا امتیازی توسط این‌جانب یا فرد دیگری در هیچ کجا ارائه نشده است و در نگارش متن پایان‌نامه شیوه‌ی نگارش مصوب دانشکده‌ی اقتصاد را به‌طور کامل رعایت نموده‌ام. چنان‌چه در هر زمان خلاف آن‌چه گواهی نموده‌ام مشاهده گردد خود را از آثار حقیقی و حقوقی ناشی از دریافت مدرک کارشناسی ارشد محروم می‌دانم و هیچ‌گونه ادعایی نخواهم داشت.

تاریخ: **امضا دانشجو:**

تقدیم به

زیباترین همراه راه ابدیت

سپاس گزاری

پیش‌تر از همه از خانواده‌ی عزیزم که همواره بی دریغ یاور و پشتیبانم بوده اند، از پدرم، مادرم که فداکارانه خودشان را وقف آینده‌ی فرزندانشان نموده‌اند، بی نهایت سپاسگزارم .
از دوستان محترمم آقایان اکبر حیدربیگی، مهرداد مددی، محمدفیاض و مهدی فیاض‌بخش به پاس توصیه های دلسوزانه ایشان نسبت به اینجانب سپاسگزارم.
از اساتید گرانقدر دست اندرکار پایان‌نامه که مهربانانه و صمیمانه دوشادوش، با صرف دقت، زمان و زحمات فراوان کار را آن‌چنان که باید سامان بخشیدند، دکتر فرزاد اسکندری راهنمای محترم پایان نامه، دکتر حمیدرضا نواب‌پور مشاور ارجمند، دکتر محمدرضا صالحی راد نماینده‌ی تحصیلات تکمیلی و دکتر خلیل شفیعی داور جلسه‌ی دفاع نهایت تشکر را دارم.
امیدوارم بتوانم از عهده ادای حق این عزیزان برآیم.

تیر ۹۱

فهرست مطالب

فهرست جدول‌ها

فهرست شکل‌ها

چکیده

برای پاسخ به بسیاری از پرسش‌های علمی، نیاز به گرداوری داده‌ها است و گرداوری داده‌ها به تنهایی پاسخ‌گوی بسیاری از پرسش‌ها نیست. بنابراین امروزه تحلیل داده‌ها در تمام شاخه‌های علمی امری ضروری به شمار می‌آید. طی چند دهه‌ی اخیر تحلیل داده‌های رده‌بندی در حوزه‌های آمار رسمی و علوم پزشکی به طور چشم‌گیری متحول شده است. با توجه به کثرت وجود این داده‌ها، بدیهی است که تحلیل و مدل‌بندی آن‌ها می‌تواند تأثیر عمده‌ای در راستای تولید دانش آماری داشته باشد. اما در تحلیل این داده‌ها با دو مشکل عمده روبرو هستیم. داده‌های همبسته و داده‌های گم‌شده. این دو مشکل فرض آماری مشاهده‌های مستقل در روش‌های رگرسیونی سنتی را نقض می‌کند. با وجود این که اکثر مطالعه‌ها برای گرداوری تمام داده‌ها طراحی می‌شوند ولی بروز گم‌شدگی در داده‌ها نیز اجتناب ناپذیر خواهد بود. مکانیسم‌های گم‌شدگی یك مسئله‌ی حائز اهمیت است زیرا خصوصیات روش‌های برخورد با گم‌شدن داده‌ها به این مکانیسم‌ها مربوط می‌شود. با در نظر گرفتن یك متغیر نشانگر برای وضعیت گم‌شدگی و توزیعی برای آن، سه مکانیسم گم‌شدگی تصادفی،کاملاً تصادفی وغیرتصادفی برای داده‌ها تعریف می‌شوند. در چنین شرایطی روش معادله‌های براوردساز تعمیم یافته معرفی می‌شود. این رویکرد یکی از مناسب ترین روش‌ها را برای تحلیل فراهم می‌کند . در این معادله‌ها تنها فرض‌هایی که اختیار می‌شوند ، فرض درباره‌ی امیدریاضی حاشیه‌ای مرتبه‌ی اول و دوم پاسخ‌ها است و هیچ فرضی درباره‌ی توزیع کامل آنها اختیار نمی‌شود . حال مسئله‌ی مورد نظر ما تعریف می‌شود. با چه روش‌هایی بر پایه GEE می‌توان به براورد پارامترها در حضور داده‌های گم‌شده پرداخت؟ از مهمترین روش‌ها در رویارویی با داده‌های گم‌شده روش جانهی است که این روش و انواع مختلف آن از جمله جانهی ساده، جانهی میانگینی و جانهی چندگانه را بیان می‌کنیم. روش دوم روش موسوم به وزن‌دهی است که در آن، سهم هر مشاهده در معادله براوردساز برابر با معکوس احتمال مشاهده شده‌ی آن است. روش بعدی استفاده از تقریب توزیع چندمتغیره‌ی نرمال است که از براورد نرمال برای پارامترهای همبستگی بهره می‌برد. روش دیگر آلگوریتم AU است که یک روش تکراری را برای حل معادله‌های براوردساز تعمیم‌یافته در حضور داده‌های گم‌شده ارائه می‌کند. در پایان تلاش خواهیم کرد با یک مطالعه شبیه‌سازی روش‌های گوناگون را مقایسه کنیم. در مجموع بر حسب اریبی و خطای استاندارد نتایج شبیه‌سازی نشان می‌دهند که روش آلگوریتم AU از سایر روش‌ها در مواجهه با داده‌های گم‌شده بهتر عمل می‌کند.

واژگان کلیدی. *معادله‌های براوردساز تعمیم یافته ، ماتریس همبستگی کاری، براوردگر استوار واریانس، داده‌های گم‌شده.*

فصل ۱
کلیات

۱-۱ مقدمه

برای پاسخ به بسیاری از پرسش‌های علمی، نیاز به گرداوری داده‌ها است. این پرسش‌های علمی در شاخه‌های مختلف علوم به شکل‌های مختلف پیش می‌آیند. برای مثال در پزشکی تأثیر چند نوع داروی مختلف بر یک بیماری مورد ارزیابی قرار می‌گیرد و در اقتصاد، تأثیر عوامل مختلف بر تورم مورد کاوش قرار می‌گیرند. در چنین کاربردهایی، نیاز به گرداوری داده‌ها امری ضروری به شمار می‌آید. پس از گرداوری داده‌ها خلاصه‌سازی آنها که نیاز به جدول‌بندی دارد، صورت می‌گیرد. پس از جدول‌بندی نیز، استخراج اطلاعات توصیفی و در پایان تحلیل این داده‌ها صورت می‌پذیرد.

تحلیل داده‌های رده بندی در حوزه‌های آمار رسمی، علوم اجتماعی و علوم پزشکی طی چند دهه‌ی اخیر به طور چشم‌گیری متحول شده است. با توجه به کثرت وجود داده‌های رده‌بندی در مطالعات مقطعی و طولی، بدیهی است که تحلیل و مدل‌بندی آن‌ها می‌تواند تأثیر عمده‌ای در راستای تولید دانش آماری داشته باشد. روش‌های مورد استفاده در تحلیل این داده‌ها از نیمه‌ی دوم قرن بیستم گسترش پیدا کرده‌اند. این روش‌ها برای تحلیل داده‌هایی که در مقیاس‌های رده‌بندی شده اندازه‌گیری می‌شوند مانند مقیاس‌های ترتیبی و اسمی اهمیت بسزایی دارند. اما در تحلیل این داده‌ها با دو مشکل عمده روبرو هستیم.

اول-داده‌های همبسته: داده‌های همبسته در مطالعات علوم اجتماعی بسیار معمول هستند. داده‌های طولی یا به صورت سلسله مراتبی ساماندهی شده، موقعیت های تحلیلی را ارائه می‌دهند که در آن داده‌ها همبسته‌اند. مثال کلاسیک در این زمینه، دانش‌آموزان دسته‌بندی شده درون کلاس‌ها در یک مدرسه هستند. این موضوع فرض آماری مشاهدات مستقل در روش‌های رگرسیون سنتی را نقض می‌کند.

دوم-داده‌های گم‌شده: به بیان ساده، مشکل داده‌های گم‌شده از تفاوت میان داده‌هایی که ما تصمیم به گرداوری داریم و داده‌هایی که گرداوری می‌کنیم ناشی می‌شود. به عنوان مثال در مطالعات طولی، این مشکل به علت عدم حضور فرد مورد آزمایش در قرار ملاقات و دست کشیدن از مشارکت در مطالعه یا خرابی تجهیزات گرداوری و در مطالعات پزشکی به علت عدم بهبودی و از دنیا رفتن فرد مورد آزمایش رخ خواهد داد. روش‌های برخورد با مسئله داده‌های گم‌شده در این گونه از مطالعات به طور پیوسته در حال پیشرفت می باشند. در این پایان نامه به بررسی یکی از مناسب ترین تکنیک‌ها برای مواجهه با این دو مشکل، یعنی روش معادله‌های براوردساز تعمیم‌یافته پرداخته

می شود.

۱-۲ مسئله‌ی مورد بررسی

هدف اولیه از تحلیل در بیشتر مدل‌های آماری، بررسی تأثیر متغیرهای کمکی (مستقل) معین بر روی متغیر پاسخ (وابسته) است. با یک مشاهده برای هر فرد، مدل‌های خطی تعمیم‌یافته برای تحلیل می توانند به کار گرفته شوند، اما با مشاهدات مکرر همبستگی بین مقادیر نیز باید در تحلیل گنجانده شود. نتایج نظری برای براوردیابی پارامترها در مدل های رگرسیونی، به ویژه برای توزیع نرمال و متغیرهای پیوسته در دسترس هستند ولی برای داده‌های رده‌بندی کمتر توسعه یافته‌اند. یکی از مشکلات در تحلیل داده‌های غیرنرمال، فقدان کلاس غنی از مدل‌ها مانند توزیع چندمتغیره‌ی نرمال برای توزیع توأم مشاهدات است. اعتبار این تحلیل‌ها بستگی به برقراری فرض‌های بیان شده در مدل‌های فوق دارد، اما این فرض‌های کلاسیک اغلب در شرایط تجربی به طور مثال در علوم پزشکی مشاهده نمی‌شوند زیرا در چنین موقعیت‌هایی بررسی‌های مکرر از یک پاسخ در زمان‌های مختلف صورت می‌گیرد.

به عنوان مثال در تحلیل داده های شمارشی، پاسخ های دودویی و ترتیبی به دلیل وجود وابستگی بین واحدهای آزمایشی و یا برقرار نبودن بسیاری از فرض‌ها همانند نرمال بودن توزیع مشاهدات، امکان استفاده از روشهای رگرسیون معمول مانند کمترین توان های دوم وجود ندارد. به دلیل گسترش و توسعه روزافزون مطالعه‌های طولی چشم پوشی از همبستگی‌های درون-واحدی در برازش مدل‌های مناسب می تواند پژوهشگر را به استنباط‌های نادرست هدایت نماید.

در چنین شرایطی رویکرد معادله‌های براوردساز تعمیم‌یافته (GEE) اولین بار توسط لیانگ و زیگر (۱۹۸۶) ارائه گردید و سپس توسط محققان دیگری توسعه یافت. این رویکرد یکی از مناسب‌ترین روشها را برای تحلیل این‌گونه از داده‌ها یعنی پاسخ‌های شمارشی، دودویی و ترتیبی فراهم می‌کند. به طور کلی می‌توان گفت این رویکرد اصلاح مدل‌های خطی تعمیم‌یافته برای داده‌های همبسته است. معادله‌های براوردساز تعمیم‌یافته می‌توانند برای حل دامنه وسیعی از مسائل آماری به کار روند و یک چارچوب متحد و هم‌شکل برای بسیاری از تکنیک های براورد آشنا شامل ماکزیمم درستنمایی، کمترین توان های دوم، شبه‌درستنمایی و براوردهای نااریب با کمترین واریانس را فراهم کنند. در این رویکرد تنها فرض هایی که اختیار می‌شوند، فرض درباره امیدریاضی حاشیه‌ای مرتبه اول و دوم پاسخ‌ها است و درباره توزیع کامل یا درستنمایی آن‌ها فرضی اختیار نمی‌شود. این رویکرد برای پارامترهای مدل رگرسیونی و واریانس آنها براوردگرهای سازگار ارائه می‌کند. در

این رویکرد، تابعی از امیدریاضی حاشیه‌ای متغیر وابسته، به صورت تابعی خطی از متغیرهای کمکی مشخص می‌شود . همچنین فرض می‌شود که واریانس، تابعی معلوم از میانگین حاشیه‌ای است . زمانی که متغیر پاسخ نرمال چندمتغیره باشد، این روش معادل با ماکزیمم درستنمایی خواهد بود.

همبستگی بین مشاهدات در این رویکرد به عنوان پارامتر مزاحم عمل می‌کند. همچنین این همبستگی با در نظر گرفتن ماتریس‌های همبستگی کاری گوناگون که ساختارهای متفاوتی دارند، مدل‌سازی می‌شود. تعیین درست ساختار همبستگی در بهبود دقت و در نتیجه کارایی ضرایب رگرسیونی مؤثر خواهد بود. اصلی ترین مزیت *GEE* نیز براورد استوار پارامترها است که حتی در صورت عدم تعیین درست ساختار همبستگی براوردهای سازگار از پارامترهای رگرسیونی را به دست می‌دهد.

اگرچه اغلب مطالعه‌ها برای گردآوری داده‌های کامل بر روی تمام افراد شرکت کننده طراحی می‌شوند، داده‌های گم‌شده به طور معمول رخ می‌دهند و باید به شکل مناسب در تحلیل وارد گردند. خیلی اوقات محققان به سادگی از رکوردهای دارای مقادیر گمشده چشم پوشی کرده و از باقی داده‌ها برای انجام تحلیل استفاده می‌کنند. این روش تحلیل ''مورد-کامل'' نامیده می‌شود. انجام تحلیل مورد-کامل با نرم افزارهای موجود آسان است اما این تحلیل ناکارآمد خواهد بود زمانی که از متغیرهای پاسخ مشاهده شده متعلق به رکوردهای گم‌شده استفاده نمی‌کند. حال مسئله معادله های براوردساز را درحضور داده های گم‌شده بررسی می‌کنیم. روشهای گوناگونی برای مواجهه با این مسئله وجود دارد:

رویکرد اول استفاده از معادله های براوردساز موزون است. در این رویکرد، سهم هر مشاهده در معادله براوردساز توسط معکوس احتمال مشاهده شدن وزن داده می‌شود.

رویکرد دوم استفاده از تقریب توزیع چندمتغیره‌ی نرمال است . این رویکرد از براورد نرمال برای پارامترهای همبستگی بهره می‌برد، یعنی تابع براوردسازی که براورد پارامترهای همبستگی را نتیجه می‌دهد، از درستنمایی نرمال چند متغیره استفاده می‌کند.

رویکرد سوم استفاده از روش‌های جانهی است. در جانهی میانگینی، متغیر گم‌شده با میانگین نمونه‌ای جایگزین می‌شود. در جانهی چندگانه، متغیر گم‌شده با یک نمونه تصادفی از مقادیر برازش شده جایگزین می‌گردد.

۱-۳ تاریخچه‌ی مسأله‌ی مورد بررسی

رویکرد GEE اولین بار توسط لیانگ و زیگر (۱۹۸۶) مطرح گردید. بسط و اصلاح های گوناگونی از این روش صورت گرفته است. پرنتیس (۱۹۸۸)، GEE را برای داده‌های دودویی همبسته بسط داد. میلر و همکاران (۱۹۹۳) رویکرد GEE را برای متغیرهای چندحالتی توسعه دادند. سپس رابینز و همکاران (۱۹۹۴) معادله‌های براوردساز موزون را مطرح کردند. ژائو و همکاران (۱۹۹۶) نشان دادند این معادله‌ها زمانی که داده‌ها گم‌شده هستند کاربردی و قابل اجراست. لاپیسیتز و همکاران (۱۹۹۹) از معادله‌های براوردساز موزون برای برخورد با مسئله داده‌های گم‌شده استفاده کردند.

چی و پایك (۱۹۹۷) روش‌های جانهی را برای معادله‌های براوردساز تعمیم یافته درنظر گرفتند زمانی که داده‌ها کاملاً تصادفی گم‌شده هستند. پایک در همان سال جانهی را برای داده‌های غیر از $MCAR$ مطرح کرد. چی و پایک در اواخر همان سال مدل معادله‌های براوردساز تعمیم‌یافته را برای پاسخ‌های دودویی زمانی که متغیرهای کمکی گم‌شده هستند ارائه دادند. لاپیسیتز و همکاران (۲۰۰۰) روش‌هایی را براساس تقریب توزیع چندمتغیره‌ی نرمال پیشنهاد دادند. لاپیسیتز و همکاران (۲۰۰۹) معادله‌های براوردساز توأم را برای داده‌های گم‌شده مطرح کردند. ژائو و همکاران (۲۰۰۹) آلگوریتم AU را پیشنهاد دادند.

۱-۴ اطلاع فیشر

فرض کنید برای پارامتر دلخواه θ، متغیر تصادفیX دارای تابع چگالی $f(x;\theta)$ باشد به طوری که دامنه متغیرX در تابع چگالی به بردار پارامتر θ بستگی نداشته باشد. با فرض مشتق‌پذیری تابع چگالی و $f(x;\theta)>0$ برای کمیت $v=\frac{\frac{\partial f(x;\theta)}{\partial x}}{f(x;\theta)}$ به سادگی می‌توان نشان داد که

$$E(v)=0$$

و نیز

$$E(v^2)=-E(v')=Var(v)$$

که در هر دو مورد امیدریاضی نسبت به تابع چگالی $f(x;\theta)$ با فرض وجود آن، محاسبه می‌شود. هر اندازه $Var(v)$بیشتر باشد،میزان تغییر نسبی چگالی X در θبیشتر می‌باشد. از این رو $Var(v)$ را مقدار اطلاع درباره پارامتر θ در متغیر تصادفی X می‌نامند و آن را با $I_X(\theta)$ نشان می‌دهند.

این نوع اطلاع که در آمار اهمیت و همچنین کاربرد زیادی دارد را برای اولین بار رونالد فیشر انگلیسی معرفی نمود و از این رو آن را اطلاع فیشر می‌نامند.

در حالت بردار پارامتر $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_k]$،ماتریس اطلاع فیشر دارای عناصر i_{jk} به شکل زیر است

$$i_{jj} = E\left(\frac{\partial LnL}{\partial \theta_j}\right)^2$$

$$i_{jk} = E\left(\frac{\partial LnL}{\partial \theta_j \theta_k}\right)^2$$

۱-۵ خانواده توزیع‌های نمایی

متغیر تصادفی Y که تابع احتمال آن در صورت گسسته بودن، یا تابع چگالی احتمال آن در صورت پیوسته بودن بستگی به پارامتر θ دارد را در نظر بگیرید. توزیع به خانواده نمایی تعلق دارد هرگاه بتوان تابع چگالی آن را به صورت زیر نوشت

$$f_y(y;\theta,\phi) = exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \qquad (۱-۵-۱)$$

که در آن θ پارامتر کانونی توزیع، b تابعی دلخواه، $a(\phi)$ تابعی از پارامتر پراکندگی ϕ و $c(y,\phi)$ تابع ثابت نرمال‌ساز است. ثابت نرمال‌ساز تابعی مستقل از θ است و تضمین می‌کند که مجموع تابع احتمال در حالت گسسته یا انتگرال آن در حالت پیوسته برابر با یک است. در حقیقت خانواده نمایی راهکاری مناسب برای محققین در مدل‌سازی متغیرهای پاسخ پیوسته و گسسته فراهم می‌کند. بسیاری از توزیع‌های معروف از جمله نرمال، پواسون، برنولی، نمایی و گاما به خانواده نمایی تعلق دارند.(هاردین و هیلب ۲۰۰۲)

براورد ماکزیمم درستنمایی پارامتر θ به کمک عبارت‌های زیر حاصل می‌شود

$$E(\frac{\partial \ell}{\partial \theta}) = 0 \qquad (۱-۵-۲)$$

و نیز

$$E\left(\frac{\partial^2 \ell}{\partial \theta^2} + (\frac{\partial \ell}{\partial \theta})^2\right) = 0 \qquad (۱-۵-۳)$$

که در آن $U = \frac{\partial \ell}{\partial \theta}$، تابع امتیاز فیشر نامیده می‌شود.

به کمک رابطه‌های (۱-۵-۲) و (۱-۵-۳) می‌توان برای هر مشاهده y_i از خانواده نمایی می‌توان عبارات زیر را برای مقادیر میانگین و واریانس به دست آورد

$$E(Y_i) = b'(\theta_i)$$

$$Var(Y_i) = b''(\theta_i)a(\phi)$$

اعضای معروف خانواده نمایی و عبارت‌های $a(\phi)$،$b(\theta)$ و θ مربوط به آن‌ها در جدول زیر آورده شده است. توجه کنید که در بعضی از توزیع‌ها مقدار پارامتر پراکندگی ثابت است.

جدول ۱-۵عبارات $a(\phi)$،$b(\theta)$ و θ برای توزیع‌های عضو خانواده نمایی

توزیع	θ	$b(\theta)$	$a(\phi)$
نرمال	μ	$-\frac{\theta^2}{2}$	σ^2
نمایی	$-\mu^-1$	$-\log(-\theta)$	1
گاما	$-\mu^-1$	$-\log(-\theta)$	α^-1
پواسون	$\log(\mu)$	e^θ	1
برنولی	$\log(\mu/(1-\mu))$	$\log(1+e^\theta)$	1
دوجمله‌ای	$\log(\mu/(n-\mu))$	$n\log(1+e^\theta)$	1
دوجمله‌ای منفی	$\log(\mu/(k+\mu))$	$-k\log(1-e^\theta)$	1

۱-۶ روش امتیازدهی فیشر

روش امتیازدهی فیشر برای حل عددی معادله‌های حاصل از تابع ماکسیمم درستنمایی به کار می‌رود. فرض کنید $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی مستقل و هم‌توزیع از تابع چگالی دوبار مشتق‌پذیر $f(x;\theta)$ باشند. برای محاسبه براورد ML بردار پارامتر θ مراحل زیر راطی می‌کنیم.
فرض می‌کنیم θ_0 نقطه‌ی شروع این روش باشد. در این حالت بسط سری تیلور تابع امتیاز $v(\theta)$ حول نقطه θ_0 به صورت زیر بیان می‌شود

$$v(\theta) \approx v(\theta_0) - \boldsymbol{J}(\theta_0)(\theta - \theta_0) \qquad (۱-۶-۱)$$

که در آن$\boldsymbol{J}(\theta_0)$ ماتریس اطلاع مشاهده‌شده در θ_0 است که به صورت زیر تعریف می‌شود

$$\boldsymbol{J}(\theta_0) = -\sum_{i=1}^{n}\left(\frac{\partial^2 l}{\partial\theta\partial\theta'}\right)$$

اکنون با قرار دادن $\theta = \theta^*$ و نیز به کمک رابطه (۱-۶-۱) می‌توان عبارت زیر را به دست آورد

$$\theta^* = \theta_0 + \boldsymbol{J}^{-1}(\theta_0)v(\theta_0)$$

بنابراین روش امتیازدهی فیشر به شکل زیر خواهد بود

$$\theta_{m+1} \approx \theta_m + \boldsymbol{J}^{-1}(\theta_m)v(\theta_m)$$

تحت شرایط نظم می توان نشان داد$\theta_m \xrightarrow{P} \theta^*$. در کاربرد$\boldsymbol{J}_X(\theta)$ توسط اطلاع فیشر که برابر با$\boldsymbol{I}_X(\theta) = E(\boldsymbol{J}(\theta))$ است، جایگزین می‌شود. در این صورت روش امتیازدهی فیشر به شکل زیر

بازنویسی می‌شود

$$\theta_{m+1} \approx \theta_m + \boldsymbol{I}_X^{-1}(\theta_m)\nu(\theta_m)$$

۱-۷ آلگوریتم نیوتون-رافسون

این آلگوریتم، تابع لگاریتم درستنمایی را در همسایگی حدس اولیه انتخاب شده، تقریب می‌زند که این تقریب توسط یک تابع انجام می‌گیرد و محل ماکسیمم این چندجمله‌ای تقریبی، به عنوان حدس دوم برای براورد ML در نظر گرفته می‌شود. سپس تابع لگاریتم درستنمایی در همسایگی نقطه دوم توسط تابع سهمی شکل دیگری تقریب زده می‌شود و حدس سوم به کمک محل ماکسیمم آن تابع به دست می‌آید و این روند ادامه می‌یابد.

برای محاسبه براورد ضرایب رگرسیونی به روش نیوتون-رافسون مشتق تابع لگاریتم درستنمایی نسبت به بردار پارامترهای رگرسیونی β به کمک قانون زنجیره‌ای در مشتق‌گیری به دست می‌آید که یک معادله به صورت زیر است

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_i} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_i}$$

و وقتی جملات مشتق جداگانه محاسبه شوند، داریم

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{y_i - b'(\theta_i)}{a(\phi)}\right)\left(\frac{1}{v(\mu_i)}\right)\left(\frac{\partial \mu}{\partial \eta}\right)_i x_{ij} \qquad (۱-۷-۱)$$

که یک معادله براورد برای محاسبه β است که به همراه تقریب خطی بسط سری تیلور زمانی که مشتق لگاریتم تابع درستنمایی وجود دارد، استفاده می‌شود.

دیدگاه عمومی در این زمینه برای براورد پارامترهای رگرسیونی β، حل معادله براورد زیر است

$$\ell'(\beta) = 0$$

که به کمک بسط سری تیلور به صورت زیر تقریب زده می‌شود

$$0 = \ell'(\beta^{(0)}) + (\beta - \beta^{(0)})\ell''(\beta^{(0)}) + \frac{1}{2}(\beta - \beta^{(0)})'\ell'''(\beta^{(0)})(\beta - \beta^{(0)}) + \dots$$

که در آن $\ell' = \frac{\partial \ell}{\partial \beta}$ و $\ell'' = \frac{\partial^2 \ell}{\partial \beta \partial \beta'}$.

برای انجام محاسبات لازم است مقدار $\beta^{(0)}$ به عنوان نقطه شروع آلگوریتم در نظر گرفته شود. در این مرحله به واسطه تقریب خطی تنها دو عبارت اول در محاسبات استفاده می‌شوند و معادله براورد بردار پارامترهای رگرسیونی به شکل ساده‌تر زیر در خواهد آمد

$$0 \approx \ell'(\beta^{(0)}) + (\beta - \beta^{(0)})\ell''(\beta^{(0)})$$