

رهیافت بیزی برای تحلیل داده‌های شمارشی در مطالعات طولی با استفاده از مدل پواسون-لگ گامای تعمیم یافته

مرضیه کمره ئی^۱، تابان باغفلکی^۲، اشکان شباک^۳، مجتبی گنجعلی^۴

تاریخ دریافت: ۹۹/۸/۲۹

تاریخ پذیرش: ۹۹/۱۲/۲۷

چکیده:

در تحلیل داده‌های طولی با پاسخ شمارشی معمولاً از مدل اثرهای تصادفی پواسون یا دو جمله‌ای منفی با فرض توزیعی نرمال برای آن‌ها استفاده می‌شود. ممکن است در برخی از کاربردها، توزیع اثرهای تصادفی نرمال نباشد. این در حالی است که بدمشخص سازی توزیع اثرهای تصادفی ممکن است باعث کاهش کارایی برآورد گرهای موجود شود. در این مقاله از توزیع لگ گامای تعمیم یافته که توزیع نرمال را به عنوان یک حالت خاص در بردارد، به عنوان فرض توزیعی اثرهای تصادفی استفاده شده است. از آنجاکه تحلیل فراوانی گرای مدل با محاسبات پیچیده مواجه است، تحلیل بیزی این مدل معرفی و برای تحلیل چندین مجموعه‌ی داده‌های واقعی استفاده می‌شود. همچنین در یک مطالعه‌ی شبیه سازی عملکرد مدل پیشنهادی مورد ارزیابی قرار گرفته است.

واژه‌های کلیدی: توزیع لگ گامای تعمیم یافته، مدل آمیخته‌ی پواسون، بیش پراکنش، مدل‌های اثرهای تصادفی، مدل دو جمله‌ای منفی چندمتغیره.

۱ مقدمه

یک چارچوب سلسله مراتبی^۵، با ارائه‌ی توزیع‌های انعطاف پذیر برای اثرهای تصادفی در مدل‌های آمیخته‌ی خطی تعمیم یافته، سعی در حل این مشکل داشته‌اند. گرچه طرح پیشنهادی لی و نلدر [۱۶] می‌تواند هر ترکیبی بین پاسخ‌های شرطی و توزیع‌های اثرهای تصادفی را شامل شود اما می‌توان گفت بیشترین کاربرد آن، در توزیع‌های مزدوج برای اثرهای تصادفی است. به هر حال با در نظر گرفتن توزیع‌های مشخص هم می‌توان همین روال را انجام داد. هاردین و هیلب [۱۰] یک فرم بسته برای توزیع حاشیه‌ای در مدل‌های پواسون با عرض از مبدأ تصادفی به دست آوردند که اثرهای تصادفی آن از توزیع لگ گاما پیروی می‌کند. مولنبرگز و همکاران [۲۱] آمیخته‌ای از اثرهای تصادفی نرمال و گاما در مدل‌های آمیخته‌ی پواسون را استفاده کردند. آن‌ها با تعمیم چارچوب کلی، اثرهای تصادفی مزدوج و نرمال را در مؤلفه‌ی سیستماتیک به منظور مدل سازی میانگین مدل‌های آمیخته‌ی خطی تعمیم یافته ترکیب کردند. همچنین، ژانگ و همکاران [۲۷] یک توزیع تعمیم یافته برای اثرهای تصادفی در مدل‌های پواسون با عرض از مبدأ تصادفی پیشنهاد کردند به طوری که مؤلفه‌ی اثر تصادفی آن از توزیع لگ گامای تعمیم یافته پیروی می‌کند. قابل اشاره است که توزیع لگ گامای تعمیم یافته توسط پرنیتس

رایج ترین مدل برای تحلیل داده‌های طولی شمارشی، مدل آمیخته‌ی خطی تعمیم یافته است. برای سادگی محاسبات در مدل‌های آمیخته‌ی خطی تعمیم یافته^۵، به دلیل غیرواقعی بودن فرض نرمال در بعضی از کاربردها، این موضوع در تحقیقات اخیر مورد بررسی قرار گرفته است. نتایج تحقیقات حاکی از آن است که اثر بدمشخص سازی توزیع‌های اثرهای تصادفی در مدل‌های آمیخته‌ی خطی باعث کاهش کارایی برآوردهای پارامترهای موجود در مدل خواهد شد. در مورد اهمیت استفاده از این فرض‌ها و حساسیت نتایج نسبت به آن‌ها اختلاف نظر وجود دارد. به منظور محافظت در مقابل بدمشخص سازی، فرض‌های توزیعی انعطاف پذیری برای اثرهای تصادفی استفاده می‌شود. لیتیره و همکاران [۱۷] بدمشخص سازی توزیع اثرهای تصادفی برای متغیرهای پاسخ در مدل لوژستیک با عرض از مبدأ تصادفی را با استفاده از روش مونت کارلو بررسی کردند. آلونسو و همکاران [۲] خانواده‌ای از آزمون‌ها جهت شناسایی بدمشخص سازی توزیع اثرهای تصادفی در مدل‌های آمیخته‌ی خطی تعمیم یافته معرفی کردند. لی و نلدر [۱۶] تحت

^۱گروه آمار، دانشگاه تربیت مدرس

^۲گروه آمار، دانشگاه تربیت مدرس (نویسنده مسئول) t.baghfalaki@modares.ac.ir

^۳پژوهشکده آمار ایران

^۴گروه آمار، دانشگاه شهید بهشتی

^۵Generalized linear model

^۶Systematic component

برای برآورد پارامترها روش‌های انتگرال‌گیری عددی وجود دارند اما در حالت کلی، مشتق‌گیری از توزیع‌های حاشیه‌ای پاسخ به شرط پارامترها، دشوار است. مدل دوجمله‌ای منفی چندمتغیره با فرض در نظر گرفتن شرایطی روی پارامترهای آن، مدل پواسون-لگ گامای تعمیم یافته را نتیجه می‌دهد. فایو و همکاران [۸] محاسبات عددی برای یافتن برآورد ماکسیمم درستمایی در مدل دوجمله‌ای منفی توسط معادلات برآوردگر تعمیم یافته را استخراج کردند. در این مقاله به بحث در روش بیزی برای برآورد پارامترهای مدل پیشنهادی آن‌ها پرداخته شده است. در این راستا با استفاده از مطالعه شبیه‌سازی، صحت و دقت مدل بررسی شده و برای تحلیل دو مجموعه داده واقعی از مدل بیزی فوق استفاده می‌شود. این مقاله به این ترتیب سازمان‌دهی شده است. در بخش دوم از این مقاله، ابتدا مروری کوتاه بر توزیع لگ گامای تعمیم یافته و دوجمله‌ای منفی چندمتغیره خواهیم داشت. سپس در بخش ۳ مدل دوجمله‌ای منفی چندمتغیره بیان می‌شود. در بخش ۴ مدل پواسون لگ گامای تعمیم یافته بحث می‌شود. در بخش پنجم به مدل بندی داده‌های طولی و خوشه‌ای شمارشی با مدل پواسون-لگ گامای تعمیم یافته و پواسون-نرمال پرداخته و در مورد برآورد پارامترهای آن به روش بیزی بحث می‌شود. بخش‌های بعد به کاربرد و شبیه‌سازی از مدل‌های توصیف شده پرداخته است. این مقاله با بحث و نتیجه‌گیری خاتمه می‌یابد.

۲ توزیع لگ گامای تعمیم یافته و دوجمله‌ای منفی چندمتغیره

۱.۲ توزیع لگ گامای تعمیم یافته

فرض کنید Y یک متغیر تصادفی است که از توزیع لگ گامای تعمیم یافته پیروی می‌کند، در این صورت گوئیم $Y \sim GLG(\mu, \sigma, \lambda)$ و تابع چگالی احتمال Y را به صورت زیر نشان می‌دهیم [۱۵].

$$f(y; \mu, \sigma, \lambda) = \begin{cases} \frac{c(\lambda)}{\sigma} \exp\left[\frac{(y-\lambda)}{\lambda\sigma} - \frac{1}{\lambda}\right] \exp\left[\frac{\lambda(y-\mu)}{\sigma}\right], & \lambda \neq 0, \\ \frac{1}{\sigma\sqrt{\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}, & \lambda = 0, \end{cases}$$

که در آن $\mu \in R$ ، $\sigma > 0$ و $\lambda \in R$. توجه کنید که μ پارامتر مکان، σ پارامتر مقیاس و λ پارامتر شکل است و همچنین $c(\lambda) = \frac{|\lambda|}{\Gamma(\lambda-1)} (\lambda-2)^{\lambda-2}$ که در آن $\Gamma(\cdot)$ تابع گاما است.

اگر $\lambda = 1$ ، توزیع مقدار کرانگین اتفاق می‌افتد. اگر $\lambda < 0$ ، تابع چگالی احتمال Y ، چوله به راست است. اگر $\lambda > 0$ ، تابع چگالی احتمال، چوله به چپ است و اگر $\lambda = 0$ ، تابع چگالی احتمال، نرمال استاندارد خواهد شد که در آن توزیع لگ گامای تعمیم یافته به صورت $Y \sim GLG(0, 1, \lambda)$ است. شکل

[۱۸] معرفی شده است. لاولس [۱۴] به عنوان یک حالت خاص، توزیع نرمال و توزیع‌های مقدار کرانگین را مطرح کرد و حالت‌های چوله به راست و چوله به چپ را در نظر گرفت. مدل‌های لگ گامای تعمیم یافته، کاربرد وسیعی در بسیاری از زمینه‌های تحلیل بقا و قابلیت اعتماد دارد. به عنوان مثال دیسیسیو [۱۶] استنباط‌هایی تقریبی برای چندک‌ها و پارامترهای مقیاس انجام داد. یانگ و باکر [۲۶] آریبی مرتبه $n^{(-1)}$ را برای پارامتری که نمونه‌های سانسور نشده در مدل‌های رگرسیونی لگ گامای تعمیم یافته را برآورد می‌کند، به دست آوردند. همچنین مشتق‌های مختلفی از لگاریتم تابع درستمایی با فرم بسته را ارائه دادند. آن [۱] یک روش رگرسیون درختی برای طبقه‌بندی کردن مجموعه‌های نامگون داده‌ها درون مدل‌های رگرسیونی لگ گامای تعمیم یافته‌ی متفاوت، پیشنهاد کرد که با هر پارامتر شکل، به طور مجزا در هر سطحی تحت سانسور تصادفی مستقل برآورد می‌شود. ماتریس‌های مناسب برای ارزیابی اثر جانبی روی پارامتر که تحت طرح‌های اختلالی متفاوت برآورد می‌شود، توسط اورتگا و همکاران [۲۲] استخراج شد. چین-تای و همکاران [۴] یک روش شرطی استنباطی برای به دست آوردن بازه‌ی اطمینان با فرض پارامتر شکل معلوم در مدل‌های رگرسیونی لگ گامای تعمیم یافته با داده‌های سانسور شده و همچنین توابع چندک‌ها و قابلیت اعتماد همراه با سانسور نوع دوم را ارائه کردند. در سال‌های اخیر، کاکس و همکاران [۵] مروری بر کاربردهای تابع خطر یک توزیع گامای تعمیم یافته در مطالعات بقای افراد مبتلا به ایدز بعد از تشخیص بالینی افراد طی دوره‌های مختلف درمانی ارائه کردند. اورتگا و همکاران [۲۳] مدل رگرسیونی لگ گامای تعمیم یافته با کسر درمانی را ارائه کردند که روی ارزیابی اثر موقعیتی تأکید می‌کند. همچنین، در تحلیل داده‌های شمارشی، توزیع لگ-چوله-نرمال پواسون با در نظر گرفتن توزیع پیشینی نرمال-لگ-چوله نرمال توسط حسن زاده و کاظمی [۹] ارائه شد. مدل گامای آمیخته با تابع ربط لگاریتمی و اثرهای تصادفی با توزیع نرمال برای داده‌های طولی توسط ریبرو و همکاران [۲۴] استفاده شد. همچنین، تحلیل مدل داده‌های توالی اسید ریبونوکلیک با استفاده از فرض توزیعی دوجمله‌ای منفی چندمتغیره برای شمارش اکسون یک ژن خاص از هر نمونه انفرادی توسط کازاکیوایز و همکاران [۱۱] انجام شد. در این مقاله، یک مدل پواسون با عرض از مبدأ تصادفی، مورد بررسی قرار گرفته که اثرهای تصادفی آن از یک توزیع لگ گامای تعمیم یافته پیروی می‌کند. این اثرهای تصادفی، بیش‌پراکنش در مدل را در نظر گرفته و همبستگی درون خوشه‌ای را مدل بندی می‌کند. مزیت توزیع لگ گامای تعمیم یافته، در انعطاف پذیری پارامترهای آن است که شکل‌های متقارن، چوله به راست و چپ را نتیجه می‌دهد. به عنوان حالت خاص این توزیع، می‌توان به توزیع نرمال اشاره کرد. از مدل پواسون-لگ گامای تعمیم یافته، با فرض یک پارامتر خاص، مدل دوجمله‌ای منفی چندمتغیره استخراج می‌شود. اگرچه

همچنین، همبستگی درون طبقه‌ای بین Y_{ij} و $Y_{i'j}$ برای $j' \neq j$ به صورت زیر بیان می‌شود:

$$Corr(Y_{ij}, Y_{i'j}) = \frac{\sqrt{\mu_{ij}\mu_{i'j}}}{\sqrt{\phi + \mu_{ij}}\sqrt{\phi + \mu_{i'j}}}$$

که این ضرایب همبستگی، همیشه مثبت هستند و برای مقادیر بزرگ ϕ ، متغیرهای تصادفی Y_{ij} در مدل دوجمله‌ای منفی چندمتغیره^۸ به طور تقریبی به عنوان مشاهدات مستقل دارای توزیع پواسون با میانگین مربوطه μ_{ij} رفتار می‌کند. در روشی دیگر، توزیع دوجمله‌ای منفی چندمتغیره را می‌توان با فرض یک توزیع لگ گامای خاص برای اثرهای تصادفی در مدل پواسون با عرض از مبدأ تصادفی استخراج کرد. محاسبات نشان خواهد داد که توزیع حاشیه‌ای Y_{ij} ، یک توزیع دوجمله‌ای منفی با میانگین و واریانس به صورت زیر

$$E(Y_{ij}) = \mu_{ij},$$

$$Var(Y_{ij}) = \mu_{ij} + \frac{\mu_{ij}^2}{\phi},$$

است که $\phi > 0$ پارامتر بیش پراکنش خواهد بود [۱۹]. همچنین می‌توان نشان داد $\mathbf{Y}_i \sim MNB(\mu_i, \phi)$ که در آن $\mu_i = (\mu_{i1}, \dots, \mu_{im_i})$ است.

۳ مدل دوجمله‌ای منفی چندمتغیره برای تحلیل داده‌های شمارشی

مدل دوجمله‌ای منفی چندمتغیره با فرض $\log \mu_{ij} = \mathbf{x}'_{ij}\beta$ در مدل (۱) تعریف شده است؛ بنابراین تابع لگاریتم درستنمایی برای مدل دوجمله‌ای منفی با فرض $\theta = (\beta', \phi)'$ عبارت است از:

$$L(\theta) = \sum_{i=1}^n \log \left\{ \frac{\Gamma(\phi + y_{i+})}{\Gamma(\phi)} \right\} - \sum_{i=1}^n \sum_{j=1}^{m_i} \log Y_{ij}! + n\phi \log \phi$$

$$- \phi \sum_{i=1}^n \log \left(\phi + \sum_{j=1}^{m_i} e^{\mathbf{x}'_{ij}\beta} \right)$$

$$+ \sum_{i=1}^n \sum_{j=1}^{m_i} Y_{ij} \log \left(\frac{e^{\mathbf{x}'_{ij}\beta}}{\sum_{j=1}^{m_i} e^{\mathbf{x}'_{ij}\beta}} \right).$$

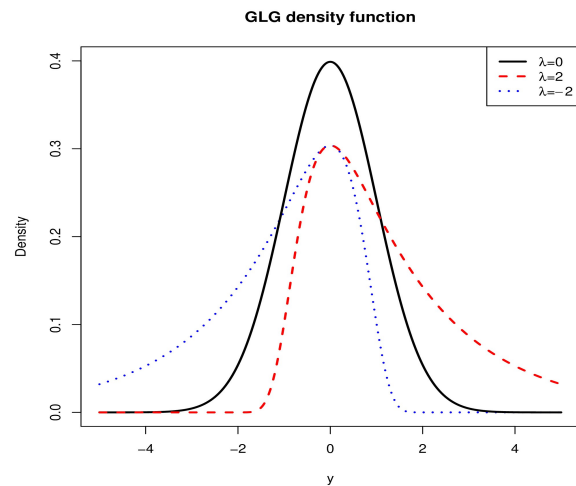
همچنین می‌توان تابع نمره و ماتریس اطلاع فشر برای θ را به دست آورد و از فرایند تکراری به منظور برآورد ماکسیمم درستنمایی استفاده کرد.

۱ تابع چگالی احتمال Y را با فرض $\mu = 0$ و $\sigma = 1$ برای مقادیر مختلف λ نشان می‌دهد. همچنین، برای $\lambda \neq 0$ گشتاور اول و دوم به صورت

$$E(Y) = \mu + \sigma \left\{ \frac{\psi(\lambda^{-2}) + \log \lambda^{-2}}{|\lambda|} \right\},$$

$$Var(Y) = \frac{\sigma^2 \psi'(\lambda^{-2})}{\lambda^2},$$

خواهند بود که در آن‌ها $\psi(\cdot)$ و $\psi'(\cdot)$ به ترتیب توابع دی گاما و تری گاما هستند.



شکل ۱. تابع چگالی توزیع لگ-گامای تعمیم یافته با فرض $\mu = 0$ و $\sigma = 1$ برای مقادیر مختلف λ

۲.۲ توزیع دوجمله‌ای منفی چندمتغیره

اگر در مدل پواسون-لگ گامای تعمیم یافته با عرض از مبدأ برای $\lambda > 0$ فرض $\sigma = \lambda$ را داشته باشیم، آنگاه تابع جرم احتمال حاشیه‌ای Y_i به صورت زیر است:

$$f_Y(\mathbf{Y}_i; \beta, \phi) = \frac{\Gamma(\phi + y_{i+})\phi^\phi \exp\left(\sum_{j=1}^{m_i} Y_{ij} \log \mu_{ij}\right)}{\left(\prod_{j=1}^{m_i} Y_{ij}!\right) \Gamma(\phi) (\phi + \mu_{i+})^{(\phi + y_{i+})}}, \quad (1)$$

که در آن $j = 1, \dots, m_i$ ، $i = 1, \dots, n$ ، $Y_{ij} = 0, 1, \dots$ که این همان تابع جرم احتمال حاشیه‌ای توزیع دوجمله‌ای منفی چندمتغیره است [۱۲] که میانگین، واریانس و کواریانس این توزیع به شرح زیر هستند:

$$E(Y_{ij}) = \mu_{ij},$$

$$Var(Y_{ij}) = \mu_{ij} + \frac{\mu_{ij}}{\phi},$$

$$Cov(Y_{ij}, Y_{i'j'}) = \frac{\mu_{ij}\mu_{i'j'}}{\phi}, (j \neq j').$$

Probability mass function^۹
Multivariate Negative Binomial (MNB)^۹

برای استخراج تابع جرم احتمال به صورت زیر عمل می‌کنیم:

$$\begin{aligned} f_Y(\mathbf{Y}_i, b_i; \beta, \phi) &= \left\{ \prod_{j=1}^{m_i} f_{Y|b}(Y_{ij}|b_i; \beta) \right\} f_b(b_i; \phi) \\ &= \prod_{j=1}^{m_i} \frac{e^{-u_{ij}} u_{ij}^{Y_{ij}}}{Y_{ij}!} \times \frac{c(\lambda)}{\sigma} \\ &\quad \exp \left\{ \frac{\lambda(b_i - \mu_{ij})}{\sigma \lambda} - \frac{1}{\lambda^\gamma} \exp \left\{ \frac{\lambda(b_i - \mu_{ij})}{\sigma} \right\} \right\} \\ &= \prod_{j=1}^{m_i} \frac{e^{-\mu_{ij} \exp(b_i)} (\mu_{ij} \exp(b_i))^{Y_{ij}}}{Y_{ij}!} \times \frac{|\lambda|}{\Gamma(\lambda - \gamma)} (\lambda - \gamma)^{\lambda - \gamma} \\ &\quad \times \frac{1}{\sigma} \exp \left\{ \frac{\lambda(b_i - \mu_{ij})}{\lambda \sigma} - \frac{1}{\lambda^\gamma} \exp \frac{\lambda(b_i - \mu_{ij})}{\sigma} \right\}. \end{aligned}$$

حال با فرض $\sigma = \lambda$ برای $\lambda > 0$ و همچنین فرض $\phi = \lambda^{-\gamma}$ داریم:

$$\begin{aligned} f_Y(\mathbf{y}_i, b_i; \beta, \phi) &= e^{-\sum_j (\mu_{ij} + \exp(b_i))} \prod_{j=1}^{m_i} \frac{(\mu_{ij} \exp(b_i))^{y_{ij}}}{y_{ij}!} \\ &\quad \times \frac{\phi^{-\gamma}}{\Gamma(\phi)} (\phi)^\phi \frac{1}{\lambda} \exp \left\{ \frac{(b_i - \mu_{ij})}{\phi} - \phi^\gamma \exp(b_i - \mu_{ij}) \right\} \\ &= \prod_{j=1}^{m_i} \frac{(\mu_{ij} \exp(b_i))^{y_{ij}}}{y_{ij}!} \times \frac{\phi^\phi}{\Gamma(\phi)} \times \phi^{-\gamma} \\ &\quad \exp \left\{ \phi(b_i - \mu_{ij}) - \phi^\gamma \exp(b_i - \mu_{ij}) \right\}. \end{aligned}$$

حال با فرض

$$\begin{aligned} \mu_{i+} &= \sum_{j=1}^{m_i} \mu_{ij}, \\ Y_{i+} &= \sum_{j=1}^{m_i} Y_{ij}, \end{aligned}$$

تابع چگالی توأم (\mathbf{y}_i, b_i) به صورت زیر نوشته می‌شود:

$$\begin{aligned} f_Y(\mathbf{y}_i, b_i; \beta, \phi) &= \frac{\phi^\phi \left(\prod_{j=1}^{m_i} \mu_{ij}^{y_{ij}} \right)}{\left(\prod_{j=1}^{m_i} Y_{ij}! \right) \Gamma(\phi)} \times \\ &\quad \exp \{ -\mu_{i+} \exp(b_i) - \phi \exp(b_i) + b_i y_{i+} + b_i \phi \}, \end{aligned}$$

با تبدیل متغیر و اینکه $\frac{db_i}{dt_i} = \frac{1}{t_i}$ ، تابع چگالی توأم (\mathbf{y}_i, t_i) را به صورت زیر می‌نویسیم:

$$\begin{aligned} f_Y(\mathbf{y}_i, t_i; b, \phi) &= \frac{\phi^\phi \left(\prod_{j=1}^{m_i} \mu_{ij}^{y_{ij}} \right)}{\left(\prod_{j=1}^{m_i} Y_{ij}! \right) \Gamma(\phi)} e^{-t_i} \left(\frac{t_i}{\mu_{i+} + \phi} \right)^{(y_{i+} + \phi)} \frac{1}{t_i} \\ &= \frac{\phi^\phi \left(\prod_{j=1}^{m_i} \mu_{ij}^{y_{ij}} \right)}{\Gamma(\phi) \left(\prod_{j=1}^{m_i} Y_{ij}! \right) (\phi + \mu_{i+})^{(y_{i+} + \phi)}} e^{-t_i y_{i+} + \phi - 1} \end{aligned}$$

بنابراین تابع جرم احتمال حاشیه‌ای \mathbf{y}_i را به این صورت استخراج می‌کنیم:

$$\begin{aligned} f_Y(y_i; b, \phi) &= \frac{\phi^\phi \left(\prod_{j=1}^{m_i} \mu_{ij}^{y_{ij}} \right)}{\Gamma(\phi) \left(\prod_{j=1}^{m_i} Y_{ij}! \right) (\phi + \mu_{i+})^{(y_{i+} + \phi)}} \times \\ &\quad \int_0^\infty e^{-t_i y_{i+} + \phi - 1} dt_i \end{aligned}$$

و از آنجایی که

$$\Gamma(\phi + y_{i+}) = \int_0^\infty e^{-t_i y_{i+} + \phi - 1} dt_i,$$

تابع جرم احتمال حاشیه‌ای \mathbf{y}_i به فرمول زیر کاهش می‌یابد که این، تابع جرم احتمال دو جمله‌ای منفی چندمتغیره است.

$$f_Y(\mathbf{y}_i; \beta, \phi) = \frac{\Gamma(\phi + y_{i+}) \phi^\phi \exp \left(\sum_{j=1}^{m_i} Y_{ij} \log \mu_{ij} \right)}{\left(\prod_{j=1}^{m_i} Y_{ij}! \right) \Gamma(\phi) (\phi + \mu_{i+})^{(y_{i+} + \phi)}}.$$

۴ مدل پواسون-لگ گامای تعمیم یافته

۱.۴ مدل پواسون-لگ گامای تعمیم یافته با

عرض از مبدأ تصادفی

همان‌طور که قبلاً اشاره شد، توزیع دو جمله‌ای منفی چندمتغیره را می‌توان با فرض یک توزیع لگ گامای خاص برای اثرهای تصادفی در مدل پواسون با عرض از مبدأ تصادفی استخراج کرد. در این قسمت با توجه به این دیدگاه به مدل بندی داده‌های شمارشی با استفاده از مدل پواسون-لگ گامای تعمیم یافته با عرض از مبدأ تصادفی می‌پردازیم. فرض کنید Y_{ij} ، برآمد زام در خوشه‌ی i ام یا پاسخ زام از آزمودنی i ام (در یک مطالعه‌ی طولی) باشد. مدل پواسون-لگ گامای تعمیم یافته با عرض از مبدأ تصادفی به صورت زیر است:

$$\begin{aligned} Y_{ij}|b_i &\stackrel{ind}{\sim} \text{Poisson}(u_{ij}) \\ u_{ij} &= \mu_{ij} \exp(b_i) \\ b_i &\stackrel{i.i.d}{\sim} \text{GLG}(\phi, \sigma, \lambda), \end{aligned} \quad (2)$$

که در آن $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})'$ و $\mu_{ij} = \exp(\mathbf{x}'_{ij} \beta)$ و $\beta = (\beta_1, \dots, \beta_p)'$ بردار پارامتر برای مقادیر $j = 1, \dots, m_i$ ، تبیینی و $i = 1, \dots, n$ است. در مدل پواسون-لگ گامای تعمیم یافته اگر $\lambda = 0$ آنگاه مدل پواسون-نرمال با عرض از مبدأ تصادفی را خواهیم داشت [۳] و [۲۰]. فرض کنید $f_{Y|b}(Y_{ij}|b_i; \beta)$ تابع جرم احتمال $Y_{ij}|b_i$ و همچنین $f_b(b_i; \sigma, \lambda)$

۲.۴ مدل پواسون-لگ گامای تعمیم یافته ییزی برای تحلیل داده های طولی شمارشی

چنانچه اشاره شد، مدل پواسون-نرمال برای در نظر گرفتن ساختار همبستگی در مطالعات طولی به خوبی قابل استفاده است؛ اما ساختار این مدل پدیده ی بیش پراکنش در مدل (که معمولاً در داده های شمارشی رخ می دهد) را در نظر نمی گیرد. لذا برای اثرهای تصادفی موجود در مدل توزیعی جایگزین خواهد شد که با وجود بیش پراکنش در مدل قادر به استفاده از این مدل باشیم. مدل پواسون-لگ گامای تعمیم یافته با عرض از مبدأ تصادفی برای تحلیل داده های شمارشی مدلی است که به جای توزیع نرمال برای اثر تصادفی b_i از توزیع لگ گامای تعمیم یافته استفاده می کند؛ بنابراین می توان ساختار سلسله مراتبی مدل پواسون-لگ گامای تعمیم یافته با عرض از مبدأ تصادفی را به صورت زیر در نظر گرفت:

$$\begin{aligned} Y_{ij}|b_i &\sim \text{Poisson}(\lambda_{ij}), \\ \log(\lambda_{ij}) &= x'_{ij}\beta + b_i, \\ b_i|\sigma, \lambda &\sim \text{GLG}(\circ, \sigma, \lambda), \end{aligned}$$

با در نظر گرفتن پیشینی نرمال چندمتغیره برای β و توزیع گامای وارون برای ابرپارامتر σ و پیشینی نرمال برای ابرپارامتر λ می توان از رهیافت ییزی برای این مدل ها استفاده کرد که به صورت

$$\begin{aligned} \beta &\sim N_p(\mu_\beta, \Sigma_\beta), \\ \sigma &\sim IG(a, b), \\ \lambda &\sim N(\mu_\lambda, \sigma_\lambda^2). \end{aligned}$$

نوشته خواهند شد؛ بنابراین با فرض $\sigma = \lambda$ اثر تصادفی در مدل پواسون-لگ گامای تعمیم یافته با عرض از مبدأ تصادفی به صورت $b_i \sim \text{GLG}(\circ, \lambda, \lambda)$ خواهد شد. در این حالت، محاسبات ساده تر شده و توزیع حاشیه ای شکل توزیع دوجمله ای منفی چندمتغیره را نتیجه می دهد. این وضعیت می تواند بیش پراکنش را نیز در مدل در نظر بگیرد. در این صورت ساختار مدل پواسون-لگ گامای تعمیم یافته با عرض از مبدأ تصادفی به صورت زیر است:

$$\begin{aligned} Y_{ij}|b_i, \beta, x_{ij} &\sim \text{Poisson}(\lambda_{ij}), \\ \log(\lambda_{ij}) &= x'_{ij}\beta + b_i, \\ b_i|\lambda &\sim \text{GLG}(\circ, \lambda, \lambda). \end{aligned}$$

تابع چگالی احتمال b_i باشد، آنگاه تابع جرم احتمال حاشیه ای $Y = (Y_1, \dots, Y_n)'$ با فرض $Y_i = (Y_{i1}, \dots, Y_{im_i})'$ به صورت

$$f_Y(Y; \beta, \sigma, \lambda) = \prod_{i=1}^n f_{Y_i}(Y_i; \beta, \sigma, \lambda), \quad (3)$$

است که در آن

$$f_{Y_i}(Y_i; \beta, \sigma, \lambda) = \int_{-\infty}^{+\infty} \left\{ \prod_{j=1}^{m_i} f_{Y_{ij}|b}(Y_{ij}|b; \beta) \right\} f_b(b_i; \sigma, \lambda) db_i. \quad (4)$$

انتگرال (۴) در حالت کلی فرم بسته ندارد؛ بنابراین، تابع لگاریتم درستنمایی برای مدل حاشیه ای وقتی $\lambda \neq \sigma$ دارای شکل زیر است:

$$\begin{aligned} L(\theta) &= \sum_{i=1}^n \log \int_{-\infty}^{+\infty} \left\{ \prod_{j=1}^{m_i} f_{Y_{ij}|b}(Y_{ij}|b; \beta) \right\} f_b(b_i; \sigma, \lambda) db_i \\ &= \left\{ n \log c(\lambda) - n \log \sigma + \sum_{i=1}^n \sum_{j=1}^{m_i} (Y_{ij} \log \mu_{ij} - \log Y_{ij}!) \right\} \\ &\quad \times \sum_{j=1}^{m_i} \log \int_{-\infty}^{+\infty} \exp \left\{ \lambda^{-1} \sigma^{-1} b_i - \lambda^{-\gamma} \exp(\sigma^{-1} \lambda b_i) \right. \\ &\quad \left. - \mu_{i+} \exp(b_i) + Y_{i+} b_i \right\} db_i. \quad (5) \end{aligned}$$

که در آن

$$\begin{aligned} \theta &= (\beta', \lambda, \sigma), \\ \mu_{i+} &= \sum_{j=1}^{m_i} \mu_{ij}, \\ Y_{i+} &= \sum_{j=1}^{m_i} Y_{ij}, \end{aligned}$$

تقریب انتگرال (۵) باید به وسیله روش های انتگرال گیری عددی همچون تقریب لاپلاس و یا روش مربع بندی گوس-هرمیت سازوار^۹ انجام شود. همچنین با فرض $\lambda = \sigma$ انتگرال (۵) دارای فرم بسته خواهد شد. به منظور تقریب این انتگرال می توان از روند *NLMIXED* در نرم افزار *SAS* استفاده کرد و به اجرای بیشینه سازی تابع لگاریتم درستنمایی تقریبی پرداخت. پیشگویی اثرهای تصادفی بر اساس روش ییزی تجربی^{۱۰} است و به صورت زیر به دست می آید:

$$\tilde{b}_i = E(b_i|y_i) = \frac{\int_{-\infty}^{+\infty} b_i f_{Y|b}(y_i|b; \beta) f_b(b_i; \sigma, \lambda) db_i}{\int_{-\infty}^{+\infty} f_{Y|b}(y_i|b; \beta) f_b(b_i; \sigma, \lambda) db_i},$$

که انتگرال فوق نیز در فرمول ییزی تجربی می تواند به کمک روش مربع بندی گوس-هرمیت سازوار تقریب زده شود.

که در آن $g(b_i|\lambda)$ به چگالی توزیع لگ گامای تعمیم یافته به صورت $GLG(\circ, \lambda, \lambda)$ اشاره می کند. چنانچه مشاهده می شود هیچ کدام از توزیع های پسینی کامل β, λ, b_i دارای فرم بسته نیستند و لذا برای انجام استنباط های بیزی در این ساختار نمونه گیری از این توزیع ها و به ویژه استفاده از الگوریتم متروپلیس-هستینگس ضروری است.

۳.۴ مدل پواسون-نرمال بیزی برای تحلیل

داده های طولی شمارشی

از مدل های رگرسیونی پواسونی به عنوان مدل های لگ-خطی پواسونی نیز نام برده می شود. این نام گذاری به دلیل تابع ربط لگاریتمی کانونی است. اگر $Y_i = (Y_{i1}, \dots, Y_{im_i})'$ به ازای $i = 1, \dots, n$ بردار مشاهدات طولی برای فرد i ام باشد، یک مدل آمیخته ی پواسونی برای داده های شمارشی طولی به صورت زیر است:

$$Y_{ij}|b_i, \beta, x_{ij}, z_{ij} \sim \text{poisson}(\lambda_{ij})$$

$$\log(\lambda_{ij}) = x'_{ij}\beta + z'_{ij}b_i,$$

حال فرض کنید بخواهیم ساختار بیزی برای این مدل را در نظر بگیریم، برای این منظور لازم است برای هر پارامتر نامعلوم توزیع پیشینی در نظر گرفته شود و همچنین فرض می کنیم توزیع های پیشینی از هم مستقل هستند. توزیع های پیشینی نرمال برای β ، نرمال چندمتغیره برای b_i و ویشارت و ارون برای واریانس در نظر گرفته می شود:

$$\beta \sim N(\mu_\beta, \Sigma_\beta),$$

$$b_i|G \sim N_q(\circ, G),$$

$$G \sim IW(\Psi, M).$$

اگر مدل (۵) تنها با عرض از مبدأ تصادفی در نظر گرفته شود در این صورت این مدل به صورت زیر کاهش می یابد:

$$Y_{ij}|b_i, \beta, x_{ij} \sim \text{poisson}(\lambda_{ij})$$

$$\log(\lambda_{ij}) = x'_{ij}\beta + b_i.$$

در ادامه تمامی توزیع های پیشینی طوری در نظر گرفته می شوند که ساختار ناآگاهی بخش را نتیجه دهند. با در نظر گرفتن توزیع پیشینی نرمال چندمتغیره

حال برای انجام محاسبات بیزی، مشخص سازی توزیع پیشینی برای پارامترهای نامعلوم ضروری است. به دلیل اینکه هیچ اطلاعات پیشینی در دسترس نیست، معمولاً پیشینی های بی اطلاع را برای پارامترها در نظر می گیرند. در این صورت می توان توزیع های پسینی کامل را با استفاده از فرمول بیز محاسبه کرد. فرض کنید $\theta = (\beta, \lambda)'$ بردار پارامترهای نامعلوم در مدل پواسون-لگ گامای تعمیم یافته باشد. اگر اعضای بردار پارامتر θ مستقل باشند، آنگاه توزیع پیشینی نرمال چندمتغیره با فرض معلوم بودن μ_β و Σ_β برای β و توزیع پیشینی گامای و ارون برای λ به صورت زیر خواهند بود:

$$\beta \sim N_p(\mu_\beta, \Sigma_\beta),$$

$$\lambda \sim IG(a, b).$$

برای انجام استنباط بیزی به توزیع پسینی کامل برای پارامترهای نامعلوم نیاز است. توزیع پسینی پارامتر β به صورت

$$\begin{aligned} \pi(\beta|y_i, b_i, x_i) &\propto f(y_i|\beta, b_i, x_i) \times \pi(\beta), \\ &\propto \left(\prod_{i=1}^n \prod_{j=1}^{m_i} \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!} \right) \\ &\quad \times \exp \left\{ -\frac{1}{\nu} (\beta - \mu_\beta)' \Sigma_\beta^{-1} (\beta - \mu_\beta) \right\}, \\ &\propto e^{-\sum_i \sum_j \lambda_{ij}} \prod_{i=1}^n \prod_{j=1}^{m_i} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \\ &\quad \times \exp \left\{ -\frac{1}{\nu} (\beta - \mu'_\beta)' \Sigma_\beta^{-1} (\beta - \mu_\beta) \right\}. \end{aligned}$$

است. برای توزیع پسینی کامل پارامتر λ در حالت $\lambda \neq \circ$ داریم:

$$\pi(\lambda|y_i, \beta, b_i)$$

$$\propto f(b_i|\lambda) \times \pi(\lambda)$$

$$\propto \frac{c(\lambda)}{\lambda} \exp \left(\frac{(b_i - \lambda)}{\lambda^2} - \frac{1}{\lambda^2} \exp \frac{\lambda(b_i - \mu)}{\lambda} \right) \times \frac{e^{-\frac{b_i}{\lambda}}}{\lambda^{a+1}},$$

$$\propto \frac{c(\lambda)}{\lambda} \exp \left(\frac{1}{\lambda^2} \{ (b_i - \lambda) - \exp(b_i - \mu) \} \right) \times \frac{e^{-\frac{b_i}{\lambda}}}{\lambda^{a+1}}.$$

در نهایت، توزیع پسینی کامل b_i عبارت است از:

$$\pi(b_i|\lambda, y_i, \beta) \propto f(y_i|b_i, \lambda) g(b_i|\lambda),$$

$$\propto \left(\prod_{i=1}^n \prod_{j=1}^{m_i} \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!} \right)$$

$$\times \frac{c(\lambda)}{\lambda} \exp \left(\frac{(b_i - \lambda)}{\lambda^2} - \frac{1}{\lambda^2} \exp \frac{\lambda(b_i - \mu)}{\lambda} \right),$$

$$\propto e^{-\sum_i \sum_j \lambda_{ij}} \prod_{i=1}^n \prod_{j=1}^{m_i} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!}$$

$$\times \frac{c(\lambda)}{\lambda} \exp \left(\frac{1}{\lambda^2} \{ (b_i - \lambda) - \exp(b_i - \mu) \} \right),$$

وجود دارد. شکل ۲ پروفایل تعداد تشنج‌های بیماران مبتلا به صرع در دو گروه مصرف‌کننده داروی پروگابید و دارونما را نشان می‌دهد که بیانگر کاهش قابل توجه تعداد تشنج‌ها در گروه مصرف‌کننده داروی پروگابید است. برای تحلیل این مجموعه داده، دیگل و همکاران [۷] مدل پواسون-نرمال با عرض از مبدأ تصادفی را که به صورت زیر است:

$$y_{ikj}|b_i \stackrel{i.i.d}{\sim} \text{Poisson}(u_{ikj})$$

$$b_i \stackrel{i.i.d}{\sim} N(\circ, \sigma^2)$$

را استفاده کردند که در آن

$$u_{i1\circ} = \exp(\alpha + b_i) + \log(t_\circ)$$

$$u_{i1j} = \exp(\alpha + \beta + b_i) + \log(t_j), i = 1, \dots, 28$$

$$u_{i2\circ} = \exp(\alpha + b_i) + \log(t_\circ)$$

$$u_{i2j} = \exp(\alpha + \beta + \delta + b_i) + \log(t_j), i = 29, \dots, 59$$

در این مدل، y_{ikj} به تعداد تشنج‌های فرد نام در گروه k ام و دوره‌ی نام اشاره می‌کند به طوری که

$$i = 1, \dots, n_k$$

$$n_2 = 31, n_1 = 28$$

$$k = 1, 2$$

$$j = \circ, 1, 2, 3, 4$$

$$t_\circ = 8, t_j = 8 + 2j$$

همچنین پارامتر β به اثر زمان اشاره می‌کند و پارامتر δ بیانگر اثر داروی پروگابید و پارامتر α بیانگر اثر دارونما است. در این مقاله به تحلیل بیزی این مدل می‌پردازیم. برای برازش مدلی مناسب به این مجموعه داده‌ها یک مدل پواسون-لگ گامای تعمیم‌یافته با عرض از مبدأ تصادفی در نظر گرفته شد. به زبانی ساده، در مدل (۳.۱)، فرض $b_i \stackrel{i.i.d}{\sim} N(\circ, \sigma)$ توسط فرض $b_i \stackrel{i.i.d}{\sim} GLG(\circ, \lambda, \lambda)$ جایگزین شد. برآورد پارامترها در حالت بیزی به کمک نرم‌افزار *OpenBUGS* مورد محاسبه قرار گرفت. نتایج تحلیل بیزی در جداول ۱ و ۲ آمده است. معمولاً به منظور مقایسه مدل‌ها از معیار اطلاع انحراف (DIC)^{۱۱} استفاده می‌شود که توسط اسپیگل هالتر [۲۵] به صورت زیر تعریف شده است:

$$\widehat{DIC} = \bar{D} + \hat{p}_D = \text{تعداد پارامترهای مؤثر مدل} + \text{انحراف}$$

که در آن p تعداد پارامترهاست، N تعداد کل مشاهدات و \hat{p}_D تعداد پارامترهای مؤثر است و به صورت $p_D = E[D(\theta)] - D(E[\theta])$ تعریف می‌شود. عبارت $D(E[\theta])$ کبیش میانگین پسینی است و با در نظر گرفتن

برای β و توزیع پیشینی نرمال برای b_i و پیشینی گامای وارون برای σ_b^2 داریم:

$$\beta \sim N_p(\mu_\beta, \Sigma_\beta),$$

$$b_i | \sigma_b^2 \sim N_q(\circ, \sigma_b^2),$$

$$\sigma_b^2 \sim IG(a, b).$$

برای به دست آوردن ساختار بیزی نیاز است که توزیع‌های پسینی کامل را با عرض از مبدأ تصادفی محاسبه کنیم. توزیع‌های کامل برای این ساختار به صورت زیر هستند:

$$\pi(\beta | \mathbf{y}_i, \lambda, x_i, b_i, \sigma_b^2) \propto f(\mathbf{y}_i, \lambda | \beta) \times \phi(\beta),$$

$$\propto e^{-\sum_i \sum_j \lambda_{ij}} \left(\prod_{i=1}^n \prod_{j=1}^m \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!} \right)$$

$$\times \exp\left(\frac{(\gamma)}{(\gamma)} (\beta - \mu)' (\beta - \mu_\beta)\right),$$

$$\pi(\mathbf{b}_i | y_i, \lambda, x_i, \sigma_b^2, \beta)$$

$$\propto f(\mathbf{y}_i | b_i) \times \phi(i | \sigma_b^2)$$

$$\propto e^{-\sum_i \sum_j \lambda_{ij}} \prod_{i=1}^n \sum_{j=1}^m \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!}$$

$$\times \frac{c(\lambda)}{\lambda} \exp\left(\frac{1}{\lambda^2} \{(b_i - \lambda) - \exp(b_i - \mu)\}\right),$$

$$\pi(\sigma_b^2 | y_i, \lambda, x_i, \mathbf{b}_i, \beta)$$

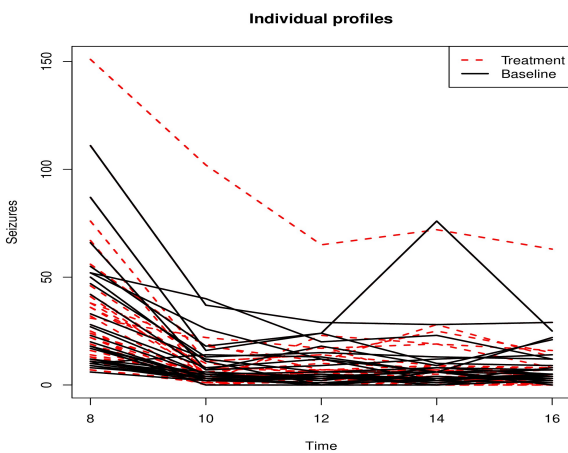
$$\propto f(b_i | \sigma_b^2) \times \phi(\sigma_b^2)$$

$$\propto \frac{c(\lambda)}{\lambda} \exp\left(\frac{(b_i - \lambda)}{\lambda \sigma_b} - \frac{1}{\lambda^2} \exp\left(\frac{\lambda(b_i - \mu)}{\sigma_b}\right)\right) \times \frac{e^{-\frac{b}{\lambda}}}{\lambda^{(a+1)}}.$$

۵ کاربرد

۱.۵ داده‌های صرع

داده‌هایی که در این بخش تحلیل می‌شوند، مربوط به ۵۹ بیمار مبتلا به صرع است که در دیگل و همکاران [۷] توسط مدل‌های طولی دیگر تحلیل شده است. این بیماران به طور تصادفی انتخاب و به دو گروه مصرف‌کننده داروی پروگابید و دارونما تقسیم شده‌اند. هدف اصلی این آزمایش، تحلیل اثر داروی شیمیایی پروگابید و اثر دارونما بر روی بیماران مبتلا به صرع است. قبل از شروع آزمایش، دوره‌ی پایه برای هر بیمار که عبارت است از سابقه‌ی تشنج در هر فرد تا قبل از شروع آزمایش، ثبت شده است. تمامی بیماران در طی ۸ هفته تحت نظر بودند و تعداد تشنج‌ها در طی دوره‌های دو هفته‌ای برای هر بیمار گزارش شده است؛ بنابراین برای هر فرد ۵ اندازه‌ی مکرر در طول زمان



شکل ۲. پروفایل تعداد تشنج‌های دو گروه مصرف‌کننده داروی پروگابید و دارونما

مقادیر میانگین نمونه‌های پسینی از پارامترهای مدل به دست می‌آید که به وسیله $\hat{D} = D \left(\frac{1}{k} \sum_{k=1}^K \theta^{(k)} \right)$ تقریب زده می‌شود. هرچه مقادیر DIC کوچک‌تر باشد به برازش بهتری از مدل اشاره می‌کند. در استفاده از روش بیزی در مدل پواسون-لگ گامای تعمیم‌یافته $DIC = 1891$ و در مدل پواسون-نرمال $DIC = 2041$ به دست آمد که برازش بهتر مدل پواسون-لگ گامای تعمیم‌یافته تأیید می‌شود. بر مبنای این نتایج زمان دارای اثری منفی است به این مفهوم که با گذشت زمان شدت تشنج‌ها کاهش می‌یابد و در ضمن داروی فوق معنی‌داری نشده است.

جدول ۱: برآورد پارامترها در مدل پواسون-لگ گامای تعمیم‌یافته برای داده‌های صرع در روش بیزی

پارامترها	میانگین	انحراف معیار	کران پایین بازه‌ی موثق	کران بالای بازه‌ی موثق	میان‌ه
α	۳۳۹۵	۰٫۸۴۷۸	۳۲۱۴	۳۳۸۵	۳۵۶۴
β	-۱۳۰۷	۰٫۴۳۹۶	-۱٫۴۱	-۱٫۲۹۹	-۱٫۲۳
δ	-۰٫۵۰۰۵	۰٫۴۹۵۴	-۰٫۱۳۱۴	-۰٫۵۵۸۳	۰٫۵۲۸۳
λ	۰٫۸۸۲۱	۰٫۰۷۶۲	۰٫۷۵۳۷	۰٫۸۷۸۱	۱٫۰۴۶
λ_1	۱٫۳۱۴	۰٫۲۲۴۱	۰٫۹۱۳۴	۱٫۲۹۷	۱٫۷۶۱

$DIC = 1891$

جدول ۲: برآورد پارامترها در مدل پواسون-نرمال برای داده‌های صرع با روش بیزی

پارامترها	میانگین	انحراف معیار	کران پایین بازه‌ی موثق	کران بالای بازه‌ی موثق	میان‌ه
α	۲٫۷۷۸	۰٫۷۷۵۷	۲٫۶۴۸	۲٫۹۲۸	۲٫۷۷۷
β	-۱٫۴۹۲	۰٫۵۰۶۷	-۱٫۵۸۵	-۱٫۳۹۸	-۱٫۴۹
δ	-۰٫۵۷۴۱	۰٫۷۳۰۴	-۰٫۱۹۲۲	۰٫۶۸۷۶	-۰٫۰۵
σ	۱٫۰۱۴	۰٫۲۰۵۲	۰٫۶۸۲۷	۱٫۴۹۷	۰٫۹۸۹۴

$DIC = 2041$

۲.۵ داده‌های سی‌دیویا

در غلظت‌های متفاوت در زمان مطالعه در شکل ۳ آمده است. لانگ و همکاران [۱۳] برای مقایسه‌ی توانمندی نیتروفن در غلظت‌های مختلف و تأثیر سم بر ارگانسیم هر یک از نرم‌تان، مدل‌های مختلف رگرسیونی پواسون با تابع ربط لگاریتمی را استفاده کردند. همچنین یک مدل رگرسیونی پواسون با اثر درجه دوم برای غلظت به‌عنوان بهترین مدل برازش توسط فابو و همکاران [۸] در نظر گرفته شد (شکل ۴)؛ اما در اکثر مطالعات صورت گرفته و از جمله این منبع مورد اشاره همبستگی بین تخم‌ریزی‌های نرم‌تان (۳ تخم‌ریزی برای هر نرم‌تن) در نظر گرفته نشده است. مدل دوجمله‌ای منفی چندمتغیره برای

در این بخش به تحلیل یک مجموعه‌ی داده با عنوان سی‌دیویا (نوعی نرم‌تن با توانایی تولیدمثل در آب‌های شیرین) پرداخته می‌شود. هدف از انجام این آزمایش، مطالعه‌ی اثر نیتروفن موجود در نوعی سم علف‌کش بر تولیدمثل غیرجنسی نرم‌تان موجود در آب‌های شیرین است. داده‌ها بیانگر تعداد تروکوفورهای متولدشده (به لارو نرم‌تان تروکوفور گفته می‌شود) در ۳ گروه ده‌تایی از نرم‌تان با ۵ غلظت متفاوت نیتروفن ۰، ۸۰، ۱۶۰، ۲۳۵ و ۳۱۰ برحسب $\frac{\text{میکروگرم}}{\text{لیتر}}$ است [۱۳]. نمایش میانگین تعداد تروکوفورهای متولدشده

برازش تعداد تروکوفورهای متولدشده می تواند به عنوان یک گزینه‌ی انتخابی در نظر گرفته شود. از این رو، مدل زیر برای برازش داده‌های سی دیویا پیشنهاد می‌شود:

که تروکوفورها از تخم بیرون می‌آیند، اشاره می‌کند (تخم‌ریزی زام برای نرم تن نام). همچنین دارای مفروضات $Day_{ij} = -2, 0, 2$ و C_{ij} به صورت زیر خواهیم بود:

$$Y_i \sim MNB(\mu_i, \phi),$$

$$\mu_{ij} = \beta_0 + \beta_1 C_{ij} + \beta_2 Day_{ij} + \beta_3 C_{ij} \times Day_{ij}$$

$$\phi = \frac{1}{\lambda^2}, i = 1, \dots, 50, j = 1, 2, 3,$$

که در آن Y_{ij} تعداد تروکوفورهای مؤنث بالغ نام در زامین تخم‌ریزی است. علاوه بر این، در این مدل C_{ij} نشان‌دهنده‌ی تغلیظ تروکوفورهایی است که در تخم‌ریزی سری زام از نرم تن نام متولدشده‌اند. همچنین Day_{ij} به روزی

$$C_{ij} = \begin{cases} 0, & i = 1, \dots, 10 \\ 80, & i = 11, \dots, 20 \\ 160, & i = 21, \dots, 30 \\ 235, & i = 31, \dots, 40 \\ 310, & i = 41, \dots, 50 \end{cases}$$

جدول ۳: برآورد پارامترها (میانگین و میانه‌ی پسینی) مدل پواسون-نرمال برای داده‌های نیتروفن با روش بی‌زی همراه با بازه‌ی موثق آن‌ها

پارامترها	میانگین	انحراف معیار	کران پایین بازه‌ی موثق	کران بالای بازه‌ی موثق	میانه
β_0	۲۵۱۱	۰/۰۹۶	۲۳۳۰	۲۷۰۸	۲۵۰۹
β_1	-۰/۰۰۴	۰/۵۵۲	-۰/۰۰۵	-۰/۰۰۳	-۰/۰۰۴
β_2	۰/۲۹۲	۰/۰۲۹	۰/۲۳۵	۰/۳۵۰	۰/۲۹۲
β_3	-۰/۰۰۱	۰/۰۱۹	-۰/۰۰۱	-۰/۰۶۴	۰/۰۰۱
σ_1	۰/۳۲۶	۰/۰۵۶	۰/۲۲۷	۰/۴۵۰	۰/۳۲۲
σ	۰/۱۰۸	۰/۰۳۷	۰/۰۵۱	۰/۱۹۹	۰/۱۰۳

$DIC = 8051$

جدول ۴: برآورد پارامترها (میانگین و میانه‌ی پسینی) در مدل پواسون-لگ گامای تعمیم‌یافته برای داده‌های نیتروفن با روش بی‌زی همراه با بازه‌ی موثق آن‌ها

پارامترها	میانگین	انحراف معیار	کران پایین بازه‌ی موثق	کران بالای بازه‌ی موثق	میانه
β_0	۲۵۳۴	۰/۰۹۵	۲۳۴۸	۲۷۲۶	۲۵۳۱
β_1	-۰/۰۰۴	۰/۵۵۹	-۰/۰۰۵	-۰/۰۰۲	-۰/۰۰۴
β_2	۰/۲۹۲	۰/۰۲۹	۰/۲۳۶	۰/۳۴۹	۰/۲۹۲
β_3	-۰/۰۰۱	۰/۰۱۹	-۰/۰۰۱	-۰/۰۸۷	-۰/۰۰۱
λ	۰/۳۰۸	۰/۰۵۸	۰/۲۰۳	۰/۴۳۳	۰/۳۰۴
λ_1	۱۲/۰۹	۴۹۷۴	۵۳۳۱	۲۵۱۳	۱۱/۲۹
ϕ	۱۱/۰۸	۳۱۲۵	۶۶۴۱	۱۶۹۴	۱۱/۱۹

$DIC = 8041$

۶ مطالعه‌ی شبیه‌سازی

در این بخش، چندین مطالعه‌ی شبیه‌سازی برای بررسی عملکرد مدل‌های پیشنهادی اجرا شده است. در این مطالعه‌ی شبیه‌سازی، ۱۰۰ نمونه با اندازه‌ی $n = 50$ از مدل دو جمله‌ای منفی چندمتغیره‌ی زیر تولید شد:

$$Y_{ij}|b_i \sim Poisson(u_{ij}), i = 1, \dots, N, j = 1, \dots, n$$

$$\log(u_{ij}) = \beta_0 + \beta_1 X_{ij} + \beta_2 t_j + \beta_3 X_{ij} + b_{itj}$$

با توجه به نتایج تحلیل‌ها در جدول‌های ۳ و ۴ و طبق معیار DIC می‌توان گفت که مجدداً در این مثال نیز مدل پواسون-لگ گامای تعمیم‌یافته با $DIC = 8041$ برازش بهتری نسبت به پواسون-نرمال با $DIC = 8051$ دارد و بر مبنای نتایج اثر Day_{ij} (روزی که تروکوفورها از تخم بیرون می‌آیند) معنی‌دار است.

$$b_i \sim GLG(0, \lambda, \lambda)$$

$(\sigma = \lambda)$ و مدل پواسون-نرمال است. این شبیه‌سازی با استفاده از بسته‌ی نرم‌افزاری *R2OpenBUGS* تحت *R* و همچنین با استفاده از روش‌های بیزی، اجرا شده است. دو زنجیر موازی با مقادیر اولیه‌ی متفاوت با ۵۰۰۰ تکرار در نظر گرفته و همچنین ۵۰۰۰ تکرار اولیه به‌عنوان دور ریز حذف شد تا استنباط پسینی بر مبنای ۳۰۰۰ نمونه‌ی بعدی صورت گیرد. همگرایی مقادیر برآورد شده با استفاده از تشخیص همگرایی بررسی و نتایج نهایی در جدول ۵ گزارش شده است.

در اینجا X_{ij} مقادیرش را از بین اعداد ۰، ۸۰، ۱۶۰، ۲۳۵ و ۳۱۰ به‌عنوان متغیر کمکی اخذ کرده و $t_j = -2, 0, 2$ به‌عنوان زمان پاسخ در نظر گرفته شده‌اند. همچنین ϕ پارامتر بیش‌پراکنش است که مقادیر ۱، ۱۰، ۱۰۰ برای آن در نظر گرفته شد. هدف از این مطالعه‌ی شبیه‌سازی، مقایسه‌ی عملکرد مدل‌های دو جمله‌ای منفی چندمتغیره (یعنی مدل پواسون-لگ گامای تعمیم‌یافته با فرض

جدول ۵: نتایج مطالعه شبیه‌سازی از مدل دو جمله‌ای منفی چندمتغیره تحت دو مدل پواسون-نرمال و دو جمله‌ای منفی چندمتغیره (مقادیری که با * نشان داده شده است نشان‌دهنده عدم محاسبه است).

مدل دو جمله‌ای منفی چندمتغیره						
<i>RMSE</i>	<i>Rel Bias</i>	<i>S.E</i>	<i>Est.</i>	مقدار واقعی	ϕ	پارامترها
۰٫۴۷۴	۰٫۰۰۵	۰٫۴۷۶	۳٫۰۱۷	۳٫۰۰۰		β_0
۰٫۱۷۰	*	۰٫۱۷۰	۰٫۰۱۳	۰٫۰۰۰	۱	β_1
۰٫۱۶۸	۰٫۰۱۷	۰٫۱۶۸	۱٫۰۱۷	۱٫۰۰۰		β_2
۰٫۱۳۳	*	۰٫۱۳۳	-۰٫۰۱۱	۰٫۰۰۰		β_3

بحث و نتیجه‌گیری

در این مقاله، مدل اثرهای تصادفی برای داده‌های شمارشی با پاسخ پواسونی مورد بحث و بررسی قرار گرفت. مدل آمیخته‌ی پواسون-نرمال تشریح و نحوه‌ی برازش مدل آمیخته‌ی پواسون با عرض از مبدأ تصادفی و با توزیع لگ گامای تعمیم‌یافته برای اثرهای تصادفی مورد مطالعه قرار گرفت. با ارائه توزیع لگ گامای تعمیم‌یافته و ایجاد انعطاف‌پذیری برای توزیع با عرض از مبدأ تصادفی در مدل‌های آمیخته‌ی پواسون، برای لحاظ کردن ساختار همبستگی روی پاسخ‌ها با وجود اثرهای تصادفی در مدل‌های آمیخته خطی تعمیم‌یافته نیاز به حل انتگرال‌های پیچیده است که با استفاده از روش‌های بیزی با در نظر گرفتن ساختار سلسله مراتبی برای استنباط آماری نیازی به حل انتگرال‌های پیچیده نخواهد بود. برای بررسی عملکرد مدل و مقایسه با مدل آمیخته‌ی پواسون-نرمال چندین مطالعه شبیه‌سازی انجام شد و به تحلیل داده‌های طولی بیماران مبتلا به صرع و همچنین داده‌های سی‌دیویا (تأثیر نیتروفن بر تخم‌ریزی نوعی نرم‌تن) پرداخته شده است. بالاخره، تعمیم شیوه‌ی مورد استفاده در این مقاله برای تحلیل پاسخ‌های دیگر با توزیع دو جمله‌ای و گاما و همچنین در نظر گرفتن برای دو یا بیش از دو اثر تصادفی می‌تواند موضوع پژوهش‌های آتی باشد.

مقادیر این جدول شامل مقادیر میانگین پسینی (*Est.*)، انحراف معیار (*S.E.*)، اریبی نسبی (*Rel.Bias*) و ریشه‌ی میانگین توان دوم خطاها (*RMSE*) هستند:

$$Rel.Bias(\theta) = \frac{1}{N} \sum_{i=1}^N \left(\frac{\hat{\theta}_i}{\theta} - 1 \right),$$

$$RMSE(\theta) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)^2}.$$

انتظار می‌رود معیارهای *Rel.Bias* و *RMSE* نزدیک به صفر باشند و همچنین با افزایش حجم نمونه یک الگوی همگرا به صفر را نشان دهند. واضح است که افزایش حجم نمونه می‌تواند معیاری مفید در کاهش اریبی و انحراف استاندارد پارامترها باشد. توجه کنید که اگر در فرمول *Rel.Bias* مقدار واقعی θ صفر باشد آنگاه $\frac{\hat{\theta}_i}{\theta}$ مقدار بی‌نهایت خواهد شد که این موضوع در جدول ۵ به صورت * نشان داده شده است. نتیجه‌ی کلی این جدول، بیانگر این مطلب است که مدل پواسون-لگ گامای تعمیم‌یافته برازش بهتری برای داده‌ها است.

۰٫۹۱۹	۰٫۱۵۰	۰٫۸۰۴	۳٫۴۵۱	۳٫۰۰۰		β_0
۰٫۳۹۳	*	۰٫۳۹۵	۰٫۰۱۸	۰٫۰۰۰	۱۰	β_1
۰٫۶۹۷	۰٫۰۹۲	۰٫۶۹۴	۱٫۰۹۲	۱٫۰۰۰		β_2
۰٫۴۰۱	*	۰٫۴۰۲	-۰٫۰۲۴	۰٫۰۰۰		β_3
۱٫۳۲۴	-۰٫۴۴۱	۰٫۰۳۸	۴٫۳۲۳	۳٫۰۰۰		β_0
۰٫۰۴۸	*	۰٫۰۴۸	-۰٫۰۰۴	۰٫۰۰۰	۱۰۰	β_1
۰٫۰۰۰۳	< -۰٫۰۰۰۱	۰٫۰۰۰۳	۰٫۹۹۹	۱٫۰۰۰		β_2
۰٫۰۰۰۱	*	۰٫۰۰۰۱	< -۰٫۰۰۰۱	۰٫۰۰۰		β_3
مدل پواسون-نرمال						
<i>RMSE</i>	<i>Rel Bias</i>	<i>S.E</i>	<i>Est.</i>	مقدار واقعی	ϕ	پارامترها
۰٫۶۹۱	-۰٫۲۰۷	۰٫۳۰۱	۲٫۳۷۶	۳٫۰۰۰		β_0
۰٫۴۰۷	*	۰٫۴۰۴	۰٫۰۶۵	۰٫۰۰۰	۱	β_1
۰٫۲۰۴	۰٫۱۸۳	۰٫۲۰۴	۱٫۰۱۸	۱٫۰۰۰		β_2
۰٫۴۱۵	*	۰٫۴۱۴	-۰٫۰۵۴	۰٫۰۰۰		β_3
۴٫۵۰۳	۱٫۴۸۴	۰٫۶۷۶	۷٫۴۵۳	۳٫۰۰۰		β_0
۰٫۵۰۶	*	۰٫۴۹۸	۰٫۱۰۲	۰٫۰۰۰	۱۰	β_1
۰٫۶۸۷	۱٫۱۳۸	۰٫۶۷۶	۱٫۱۳۸	۱٫۰۰۰		β_2
۰٫۵۰۴	*	۰٫۴۹۶	-۰٫۱۰۱	۰٫۰۰۰		β_3
۹٫۲۱۰	۳٫۰۷۰	۰٫۰۱۵	۱۲٫۲۱۰	۳٫۰۰۰		β_0
۰٫۰۰۰۴	*	۰٫۰۰۰۴	< -۰٫۰۰۰۱	۰٫۰۰۰	۱۰۰	β_1
۰٫۰۰۰۳	< -۰٫۰۰۰۱	۰٫۰۰۰۳	۰٫۹۹۹	۱٫۰۰۰		β_2
۰٫۰۰۰۱	*	۰٫۰۰۰۱	< -۰٫۰۰۰۱	۰٫۰۰۰		β_3

مراجع

- [1] Ahn, H. (1996), Log-Gamma Regression Modeling through Regression Trees, *Communications in Statistics, Theory and Methods*, **25**, 295-311.
- [2] Alonso, A., Litiere, S. and Molenberghs, G. (2008), A Family of Tests to Detect Misspecification in the Random-Effects Structure of Generalized Linear Mixed Models, *Computational Statistics and Data Analysis*, **52**, 4474-4486.
- [3] Breslow, N. E., Clayton, D. G. (1993), Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, **88**, 9-25.
- [4] Chien-Tai, L., Wu, S. j. S. and Balakrishnan, N. (2004), Interval Estimation of Parameters Log-Gamma Distribution Based on Progressively Censored data, *Communications in Statistics, Theory and Methods*, **33**, 2595-2626.
- [5] Cox, C., Chu, H., Shneider, M. F., Munoz, A. (2007), Parametric Survival Analysis and Taxonomy of Hazard Functions for the Generalized Gamma Distributions, *Statistics in Medicine*, **26**, 4352-4374.

- [6] Diccio. T. J. (1987), Approximate inference for the generalized gamma distributions. *Technometrics*, **29**, 33-40.
- [7] Diggle, P. J., Heagerty, P. J., Liang, k. y and Zeger, S., L. (2002), *Analysis of Longitudinal Data*, 2ed. Oxford University Press.
- [8] Fabio. L.C., Gilberto. A. Paula and Mario de Castro. (2012), A Poisson Mixed Model with Non-Normal Random Effect Distribution, *Computational Statistics and Data Analysis*, **56**, 1499-1510.
- [9] Hassanzadeh, F. and Kazemi, I. (2016) Analysis of over-dispersed count data with extra zeros using the Poisson log-skew-normal distribution, *Journal of Statistical Computation and Simulation*, **86(13)**, 2644-2662.
- [10] Hardin, J. W., Hilbe, J. M. (2003), *Generalized Estimating Equations*, Chapman and Hall, Boca Raton.
- [11] Kazakiewicz D, Claesen J, Górczak K, Plewczynski D, Burzykowski T. (2019), A Multivariate Negative-Binomial Model with Random Effects for Differential Gene-Expression Analysis of Correlated mRNA Sequencing Data. *J Comput Biol.*, **26(12)**, 1339-1348.
- [12] Johnson, N.L., Kotz, S. Balakrishnan, N. (1997), *Discrete Multivariate Distributions*, Wiley, New York.5-186.
- [13] Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L and Greenhouse, J. s (1994), *Case Studies in Biometry*, Wiley, New York.
- [14] Lawless, J. F. (1987), Negative Binomial and Mixed Poisson Regression. *Canadian Journal of Statistics*, **15**, 209-225.
- [15] Lawless, J. F. (2002), *Statistical Models and Methods for Lifetime Data*, second ed, Willey, New York.
- [16] Lee, Y., Nelder, J. A. (1996), Hierarchical Generalized Linear Models, *Journal of the Royal Statistical Society. Series B*, **58**, 619-675.
- [17] Litiere, S., Alonso, A., Molenberghs, G. (2008), The Impact of A Misspecified Random-Effects Distribution on the Estimation and the Performance of Inferential Procedures in Generalized Linear Mixed Models, *Statistics in Medicine*, **27**, 3125-3144.
- [18] Prentice, R. (1974), Log-Gamma Model and its Maximum Likelihood Estimation. *Biometrika*, **61**, 539-544.
- [19] McCullagh, P., Nelder, J. A. (1989), *Generalized Linear Mixed Models*, Second ed, Chapman and Hall, London.
- [20] McCullagh, C. E., Searle, S. R. (2001), *Generalized Linear Models*, Wiley, New York.
- [21] Molenberghs, G., Verbeke, G., Demetrio, C. G. B. (2010), A Family of Generalized Linear Models for Repeated Measures with Normal and Conjugate Random Effects, *Statistical Science*, **25**, 325-347.
- [22] Ortega, E.M.M., Bolfarine, H and Paula, G. (2003), Influence Diagnostics in Generalized Log-Gamma Regression Models, *Computational Statistics and Data Analysis*, **42**, 165-186.
- [23] Ortega, E.M.M., Cancho, V.G., Paula, G.A. (2009), Generalized Log-Gamma Regression Models with Cure Fraction. *Lifetime Data Analysis*, **15**, 79-106.
- [24] Ribeiro, M. H. D. M., Santiago, A. N., Oliveira, R. M. W. de, Milani, H., Previdelli, I. (2019). Longitudinal modeling using log-gamma mixed model: case of memory deterioration after chronic cerebral hypoperfusion associated with diabetes in rats. *Acta Scientiarum. Technology*, **41(1)**, e35789.

- [25] Spiegelhalter, D., Best, N., Carlin, B and van der Linde. (2002), Bayesian Measures of Model Complexity and Fit, *Journal of the Royal Statistical Society; Series B*, **64**, 583-639.
- [26] Young, D. H., Bakir, S. T. (1987), Bias Correction for a Generalized Log-Gamma Regression Model. *Technometrics*, **29**, 83-191.
- [27] Zhang, P., Song, P and Green, T. (2008), Efficient Estimation for Patient-Specific Rates of Disease Progression Using Non-Normal Linear Mixed Models. *Biometrics*, **64**, 29-38.