فصلنامه پژوهش‌های اقتصادی ایران/ سال بیست‌ویکم/ شماره ۶۶/ بهار ۱۳۹۵/ صفحات ۱۰۱-۸۵

برآورد بیزی همبستگی پارامترهای دو متغیر تصادفی پواسون: کاربردی برای تصمیم‌گیری بنگاه

فرزاد اسکندری[۱]

تاریخ ارسال: ۱۳۹۴/۳/۲۲ تاریخ پذیرش: ۱۳۹۵/۴/۶

چکیده

در این مطالعه براساس مدل‌های خطی تعمیم‌یافته بیز، به بررسی همبستگی پارامترهای دو توزیع پواسون پرداخته شده است. با توجه به نبود فرم بسته توزیع‌های پسین، آمار بیزی سلسله مراتبی با استفاده از الگوریتم متروپلیس- هستینگز برای بررسی همبستگی پارامترها در دو توزیع پواسون ارایه می شود. در این رابطه، بیشترین احتمال ناحیه پسین برای ضرایب متغیرهای کمکی در مدل محاسبه شده است. با استفاده از معیار اطلاع انحراف بیزی نیز نشان داده شده است مدل پواسون- لگ‌نرمال بهتر از مدل پواسون–گاما می‌تواند همبستگی بین پارامترهای دو توزیع پواسون را ارزیابی کند. در پایان، روش پیشنهادی برای داده‌های شبیه‌سازی شده بانک تجارت مورد استفاده قرار گرفته است.

واژگان کلیدی: توزیع دومتغیری گسسته، مدل پواسون– لگ‌نرمال، مدل پواسون– گاما، روش بیزی سلسله مراتبی، الگوریتم متروپلیس- هستینگز.

طبقه‌بندی JEL: *C11، C15*.

۱- دانشیار گروه آمار دانشگاه علامه طباطبائی، پست الکترونیکی: ffeskandari@yahoo.com.

۱- مقدمه

داده‌های موجود در یک جامعه آماری را می‌توان براساس نوع متغیرهای کمی یا کیفی و ارتباطی که بین آنان به‌وجود می‌آید، به سامانه‌های پیچیده‌ای تشبیه کرد که به‌طور دایم در حال تغییر هستند. بسیاری از تغییرات که در درون این سامانه‌های پیچیده به‌طور خودکار انجام می‌شود، نه برای ما اهمیت دارد و نه اینکه تغییرات آن تأثیر مهمی بر رفتار و زندگی روزمره ما می‌گذارد، اما گاهی اوقات تغییرات درون‌سامانه‌ای پیچیده در جوامع آماری می‌تواند به نوعی برای ما مهم باشد؛ برای مثال، مشتریان متقاضی دریافت وام از بانک‌ها براساس تعداد دفعاتی که وام دریافت می‌کنند و تعداد دفعات جریمه دیرکرد آن متغیرهایی هستند که به نوعی بر یکدیگر تأثیرگذارند و یکی از انواع سامانه‌های پیچیده را ایجاد می‌کنند.

بررسی تحلیلی و مدل‌سازی چنین سامانه‌های پیچیده چندمتغیره‌ای مبتنی‌بر یک مدل تصادفی پویای توانمند که در آن ترکیبی از عوامل مؤثر اندازه‌پذیر و غیراندازه‌پذیر وجود دارند، بسیار اهمیت دارد و مطالعه آن توسط یک آمارشناس خبره یا یک اقتصاددان باتجربه به تنهایی کار مناسبی نخواهد بود و به معرفی یک گروه از چندین متخصص مختلف نیاز دارد. در این‌گونه مطالعات آنچه می‌تواند برای یک آمارشناس مورد توجه قرار گیرد، شناسایی ارتباط متغیرهایی است که در این موضوع دخالت دارند. لازم است پس از شناسایی آنها و مطالعه دقیق متغیرها با ارایه مدل مناسب آماری که وضعیت پیچیده مطرح شده را مطابق با آنچه هست تفسیر کند، شاخص‌های اندازه‌پذیر و غیراندازه‌پذیر مورد نیاز را به‌طور دقیق مطالعه کند. یکی از این متغیرها که به‌طور معمول در بسیاری از مطالعات نادیده گرفته می‌شود، وابستگی پنهان یا همبستگی پنهان بین پارامترهای مربوط به توزیع‌های دو متغیر تصادفی در یک جامعه آماری است. این موضوع زمانی می‌تواند مهم باشد که توزیع هر دو متغیر از نوع گسسته یا شمارشی باشد. در این حالت، دیگر نمی‌توان در ادبیات جاری شاخصی را برای آن مطرح کرد. شاید به سبب این موضوع است که به‌طور معمول در مورد آن سخنی به میان نمی‌آید. در این مقاله، سعی می‌شود با استفاده از مدل‌های تعمیم‌یافته خطی و به شیوه استنباط بیزی، در حالتی که توزیع‌های دومتغیری مورد

بررسی پواسون دومتغیری باشند، برآوردیابی پارامترهای مربوط به همبستگی پنهان دو پارامتر توزیع‌های پواسون را به‌دست آوریم. همچنین علاوه بر بررسی خواص بهینه آنها، موضوع را برای داده‌های واقعی مطالعه می‌کنیم.

۲- ساختار مدل توزیع دومتغیری پواسون

استفاده از یک مدل آماری مناسب برای تحلیل داده‌های شمارشی چندمتغیری و بررسی همبستگی پنهان بین پارامترهای آنها، از اهداف این مقاله است. مطالعات زیادی در این باره شده است، اما برای مثال، می‌توان از کارهای بین‌المللی به خفری و همکاران (۲۰۰۸)، اشاره کرد. در این مطالعه، علاوه بر معرفی توزیع پواسون دومتغیری، به مفهوم همبستگی دو متغیر نیز اشاره شده است. در این بررسی، داده‌های پزشکی که دارای ساختار پواسون دومتغیری بودند، مورد استفاده و تحلیل قرار گرفتند. برای داده‌های ترتیبی می‌توان به کاظم‌نژاد و همکاران (۲۰۱۰)، اشاره کرد. در این مقاله نیز براساس داده‌های ترتیبی، ساختار منطبق برای متغیرها از توزیع دومتغیری پواسون تبعیت کرد. از کارهای دیگر می‌توان به اسکندری و نقی‌زاده (۱۳۸۴)، انتی‌زوفراس و همکاران (۲۰۰۵)، کارلیس و انتی‌زوفراس (۲۰۰۵) و راموز و اسکندری (۲۰۱۶)، اشاره کرد. در تمام این کارها برای داده‌های شمارشی به صورت بیزی تأثیر عوامل کمکی بر پاسخ، در حالتی که توزیع پاسخ پواسون یا چندجمله‌ای است براساس روش‌های مونت‌کارلوی زنجیره مارکوف مورد بررسی آماری قرار گرفته است. البته به کارهای قدیمی‌تر مانند کارلیس و کارلین (۱۹۹۶) نیز می‌توان اشاره کرد، اما در این کارها از روش‌های ساده مونت‌کارلو برای بررسی داده‌ها استفاده شده است. این پژوهش‌ها به بیان و استفاده از مباحث مربوط به مدل‌های خطی تعمیم‌یافته و مقایسه کارآیی مدل‌های پیشنهادی به دو شیوه فراوانی‌گرا و بیزی پرداخته‌اند. یکی از مدل‌های خطی تعمیم‌یافته که کاربرد زیادی دارد، مدل رگرسیون پواسون است. رگرسیون پواسون در واقع، ساده‌ترین و اساسی‌ترین مدل برای تحلیل داده‌های شمارشی است. گاهی اوقات دو متغیر تصادفی پواسون در مدل وجود دارند و یک همبستگی پنهان بین آنها ایجاد شده است و این

همبستگی در نظر گرفته نمی‌شود، باید به دنبال آن باشیم تا به‌وسیله یک روش مناسب و کارآمد، این میزان همبستگی را اندازه بگیریم. بررسی و اندازه‌گیری این همبستگی در کاربرد بسیار مهم است. باید توجه کرد که برابر بودن میانگین و واریانس متغیرهای وابسته پواسونی، شرط اصلی استفاده از آن است، حال آنکه برقرار نبودن این شرط به بیش‌پراکندگی در مدل منجر می‌شود. به‌طور عموم برای غلبه بر این مشکل، استفاده از مدل‌های آمیخته با استفاده از روش‌های متغیر پنهان پیشنهاد شده است. از جمله این روش‌ها، استفاده از مدل دوجمله‌ای منفی است که به صورت استاندارد برای تحلیل داده‌های شمارشی مورد استفاده قرار می‌گیرد. یکی دیگر از روش‌های رایج برای غلبه بر مشکل بیش‌پراکندگی، استفاده از توزیع‌های آمیخته پواسون- گاما و پواسون- لگ‌نرمال است. در این دو مدل، توزیع داده‌ها پواسون در نظر گرفته می‌شود، به این ترتیب که ۱- متغیرهای پنهان مستقل در نظر گرفته می‌شوند و ۲- به صورت اثرهای تصادفی وارد مدل می‌شوند. راه‌های برآورد پارامترهای مدل و استنباط در این شرایط در مطالعات اسکندری و نقی‌زاده (۱۳۸۴) و راموز و اسکندری (۱۳۸۹) مورد نقد و بررسی قرار گرفته است. در مدل‌های ارایه شده به‌راحتی ثابت می‌شود که توزیع حاشیه‌ای دوجمله‌ای منفی برابر با توزیع پواسون- گاما و پواسون- لگ‌نرمال است. همچنین با در نظرگرفتن این توزیع‌های آمیخته، بیش‌پراکندگی در مدل، تحت کنترل قرار خواهد گرفت. توزیع پواسون دومتغیره توسط جانسون و کانز (۱۹۶۹) ارایه شد که در آن، به ترتیب نرخ متغیر اول، نرخ متغیر دوم و همبستگی بین پارامترهای دو توزیع متغیر اول و دوم بیان می‌شود. همین‌طور توزیع‌های پواسون دومتغیره‌ای توسط کوچرلاکوتا و کوچرلاکوتا (۲۰۰۱) ارایه شد. در تمام توزیع‌های ارایه شده ضریب همبستگی محدود به اعداد مثبت است. در این مطالعه، از مدل ترکیبی پواسون– نرمال و به شیوه بیز استفاده می‌شود. ساختار آن به شکل زیر است:

فرض کنید، مشاهده Y_{ij} برای $i,j = 1,2,\dots,n$ مقدار پاسخ‌های مختلف باشد، به‌طوری که بین پاسخ‌ها همبستگی پنهان وجود داشته باشد و این همبستگی برای $i\#k$، $cov(Y_{ij},Y_{kl})\#b_{ij}$ و برای $i=k$، $cov(Y_{ij},Y_{kl})=b_{ij}$ شود.

همچنین فرض کنید، مشاهده Y_{ij} متغیر تصادفی پواسون با پارامتر λ_{ij} باشد، یعنی:

$$Y_{ij}|x_{ij}, b_{ij} \sim Pois(\lambda_{ij}) \quad (۱)$$

که در آن، پارامتر λ_{ij} به صورت زیر تعریف می‌شود.

$$\lambda_{ij} = exp(x_{ij}\beta_j + b_{ij}) \quad (۲)$$

براساس تعریف، همبستگی پنهان بین مشاهده Y_{i1} و مشاهده Y_{i2} در رده iام به‌وسیله همبستگی بین مؤلفه‌های ناهمگن غیرمشاهده‌ای b_1 و b_2 تعریف می‌شود که در اینجا خواهیم داشت:

$$b_i = (b_{i1}, b_{i2}) \sim N_2(0, \Sigma) \quad (۳)$$

در اینجا فرض می‌کنیم، ماتریس Σ مجهول و دارای توزیع پیشین ویشارت است. در این حالت، مدل یادشده به مدل پواسون-گامای دومتغیره تغییر خواهد یافت.

مشاهده Y_1 دارای توزیع دومتغیره لگ‌نرمال و مشاهده Y_1 دارای توزیع آمیخته پواسون – لگ‌نرمال است. آیتچیسون و هو (۱۹۸۹)، نشان دادند که میانگین و واریانس و کوواریانس توزیع شمارشی فوق به صورت زیر است:

$$E(Y_{ij}|\beta_j, \Sigma) = exp\left(\acute{x}_{ij}\beta_j + \frac{1}{2}\sigma_{ij}\right) = \lambda_{ij}$$
$$var(Y_{ij}|\beta_j, \Sigma) = \lambda_{ij} + \lambda_{ij}^2(exp(\sigma_{ij}) - 1)$$
$$cov(Y_{ij}, Y_{ij}|\beta_{ij}, \beta_{ij}, \Sigma) = \lambda_{ij}\lambda_{ik}exp(\sigma_{jk}) - 1$$

σ_{ii} و σ_{ij}، مؤلفه‌های ماتریس Σ هستند. همان‌طور که ملاحظه می‌شود، مدل پواسون - لگ‌نرمال همبستگی‌های مثبت و منفی را پوشش می‌دهد، زیرا مقدار کوواریانس به‌دست آمده متعلق به مجموعه اعداد صحیح است. در این حالت، تابع چگالی احتمال بردار شمارشی Y_i به صورت زیر است:

$$f(y_{i1}, y_{i2}|x_i, \Sigma) = \int \prod_{j=1}^{2} f(y_{ij}|\beta_j, b_{ij})\phi(b_j|0, \Sigma) \; db_j \quad (۴)$$

تابع لگاریتم درست‌نمایی غیرشرطی مدل پواسون – لگ‌نرمال دومتغیری به صورت زیر است:

$$\log(L) = \sum_{i=1}^{m}\sum_{j=1}^{2} \log P(y_{ij}|x_i)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{2} \log\left\{\frac{exp[-exp(\delta_{ij}b_{ij})\lambda_{ij}][exp(\delta_{ij}b_{ij})\lambda_{ij}]^{y_{ij}}}{y_{ij}!}\phi(b_i)\, db_i\right\}$$

انتگرال‌های به‌دست آمده در تابع لگاریتم درست‌نمایی دارای ساختار معین و بسته نیست. ازاین‌رو، به‌عنوان یک راه‌حل از طریق شبیه‌سازی می‌توان رابطه زیر را نوشت:

$$\log(L_n) = \sum_{i=1}^{m}\sum_{j=1}^{2} \log\left(\frac{1}{M}\sum_{m=1}^{M} P(y_{ij}|x_i, b_{ijm})\right)$$

که در آن، b_{ijm} متغیری تصادفی است و به‌عنوان پیشنهاد می‌توان فرض کرد، دارای توزیع نرمال استاندارد است. از آنجا که محاسبه و برآورد پارامترها دارای محاسبات سنگین است، ازاین‌رو، در این مرحله می‌خواهیم براساس روش بیز به تحلیل پارامتر همبستگی پنهان اقدام کنیم. در ابتدا با توجه به رابطه (۴) توزیع‌های پیشین ویشارت و نرمال چندمتغیره در مدل بیز سلسله مراتبی برای دو پارامتر Σ و β به‌طور مستقل در دو سطح زیر می‌تواند مورد توجه قرار گیرد. در نتیجه، براساس (۳) می‌نویسیم

$$\beta \sim N_k(\beta_\beta, V'_\beta)$$
$$\Sigma^{-1} \sim Wishart(\nu_\Sigma, V_\Sigma)$$

و برای توزیع پیشین اولیه داریم:

$$\beta_\beta \sim N_k(\beta_0, U_\beta)$$
$$V_\beta^{-1} \sim Wishart(\nu_{0\beta}, R_0)$$
$$V_\Sigma^{-1} \sim Wishart(\nu_{0\Sigma}, \Sigma_0)$$

در اینجا پارامترهای اولیه معلوم در نظر گرفته می‌شود. همچنین در این حالت فرض می‌کنیم توزیع متغیرهای فوق از یکدیگر مستقل هستند.

۳- تعیین توزیع‌های پسین

با در نظر گرفتن ساختار مدل و توزیع‌های پیشین پارامترها، می‌توان توزیع پسین پارامترها را به صورت زیر محاسبه کرد:

$$\prod f(\beta_j, \beta_\beta, V_\beta, b_i, \Sigma|y_i, X) \propto f(V_\Sigma^{-1}) f(\beta_j|\beta_\beta, V_\beta) f(\beta_\beta) f(V_\beta^{-1})$$

با توجه به تابع درست‌نمایی، توزیع پسین به صورت زیر خواهد بود.

$$\prod f(\beta_j, \beta_\beta, V_\beta, b_i, \Sigma|y_i, X) \propto \prod_{i=1}^{n}\prod_{j=1}^{2} exp\left(-\exp(X'_{ij}\beta_j + b_{ij})\right)\left(\exp(X'_{ij}\beta_j + b_{ij})\right)^{y_{ij}}$$

همان‌طور که ملاحظه می‌شود، توزیع پسین دارای فرم بسته نیست، در نتیجه، برای محاسبه برآورد پارامترهای مدل به روش‌های نمونه‌گیری با استفاده از مونت‌کارلوی زنجیره مارکوفی مانند الگوریتم گیبس و در حالت پیشرفته‌تر الگوریتم متروپلیس- هستینگز برای تقریب توزیع پسین نیاز است. باید توجه کرد، گاهی اوقات، نمونه‌گیری از بعضی از توزیع‌های پیچیده کاری بسیار مشکل است. در این باره استفاده از روش‌های مونت‌کارلوی زنجیره مارکوفی می‌تواند مشکل را حل کند. برای این منظور فرض کنید، می‌خواهیم از توزیع:

$$P(x) = \frac{f(x)}{k}$$

نمونه‌ای تصادفی را برگزینیم، به‌طوری که در آن، ثابت نرمال‌ساز k ممکن است نامعلوم و محاسبه آن مشکل باشد. الگوریتم متروپلیس - هستینگز به صورت زیر، یک نمونه تصادفی از این توزیع تولید می‌کند:

الف- با مقدار اولیه x_0 که در $f(x_0) \geq 0$ صدق کند، شروع می‌کنیم.

ب- با استفاده از مقدار x، نقطه کاندید x^* را از توزیع پیشنهادی $q(x_1, x_2)$ استخراج می‌کنیم. $q(x_1, x_2)$ احتمال به‌دست آمدن مقدار x_2 است، به‌شرط آنکه مقدار x_1 را داشته باشیم. تنها محدودیتی که روی این توزیع در الگوریتم متروپلیس وجود دارد، این است که باید متقارن باشد.

یعنی:

$$q(x_1, x_2) = q(x_2, x_1)$$

ج- با داشتن نقطه کاندید x^*، نسبت چگالی در نقطه کاندید x^* به نقطه اخیر، یعنی x_{t-1} را محاسبه می‌کنیم و x براساس آن محاسبه می‌شود.

$$\alpha = \min\left\{\frac{f(x^*)q(x^*, x_{t-1})}{f(x_{t-1})q(x_{t-1}, x^*)}, 1\right\}$$

د- اگر $x>1$ باشد، x^* را می‌پذیریم، یعنی قرار می‌دهیم $x_t = x^*$. اگر $x>1$ باشد، با احتمال α نقطه x^* را قبول می‌کنیم. در اینجا لازم است از یک روش تصادفی برنولی با احتمال α استفاده کنیم.

به این ترتیب یک زنجیره مارکوف $(x_0,...,x_n)$ تولید می‌شود که احتمال انتقال آن از x_t به x_{t+1} تنها به x_t بستگی دارد و به مقادیر x_{t-1} , ... , x_0 بستگی ندارد. بردار $(x_{k+1},...,x_{k+n})$ نمونه‌ای از $q(x)$ خواهد بود.

٤- استفاده از روش‌های شبیه‌سازی مونت‌کارلوی زنجیره مارکوفی برای برآورد پارامترها

در قسمت قبل ملاحظه شد که توزیع پسین به‌دست آمده دارای ساختار مشخص و معلومی نیست. بنابراین، برای انجام تحلیل بیزی و به‌دست آوردن برآورد پارامترها، نمی‌توان به روش معمول به تحلیل داده‌ها اقدام کرد. در این حالت، باید با استفاده از روش‌های شبیه‌سازی، از جمله روش مونت‌کارلوی زنجیره مارکوفی به نمونه‌گیری از توزیع‌های پسین پیچیده بپردازیم. برای بالا بردن دقت و کنترل خطا، لازم است از طریق نمونه‌گیری گیبس که حالت خاص نمونه‌گیری متروپلیس- هستینگز است و در حالت پیشرفته‌تر از الگوریتم متروپلیس- هستینگز به نمونه‌گیری و سپس، برآورد پارامترها اقدام کنیم. در نمونه‌گیری گیبس از توزیع‌های تک‌متغیره نمونه‌گیری انجام می‌شود. برای انجام این نمونه‌گیری ابتدا لازم است توزیع شرطی هر یک از پارامترها به‌شرط سایر پارامترها و داده‌ها (توزیع پسین هر یک از پارامترها به شرط سایر پارامترها) که اصطلاحاً به آنها توزیع شرطی گفته می‌شود، تعیین شود. در مرحله بعد، در هر تکرار از هر یک از توزیع‌های

شرطی کامل یک نمونه می‌گیریم. اگر این توزیع تمام شرطی دارای ساختار مشخصی باشد، این نمونه به‌طور مستقیم از خود آن توزیع گرفته می‌شود و اگر توزیع شرطی کامل دارای فرم مشخص و بسته‌ای نباشد، از الگوریتم متروپلیس - هستینگز استفاده می‌کنیم. به این ترتیب نمونه‌ای که به این طریق بعد از تکرار به‌دست می‌آوریم نمونه‌ای از توزیع پسین توأم خواهد بود. بنابراین، در این قسمت، ابتدا توزیع شرطی هر یک از پارامترها برای مدل‌های معرفی شده در بخش قبل به‌دست می‌آید و سپس، در ادامه، شیوه نمونه‌گیری از هر یک از این توزیع‌های تمام شرطی بیان می‌شود.

٤-١-توزیع شرطی کامل b_i

تابع چگالی پسین b_i یکی از توزیع‌های مهم بوده که در اینجا مورد توجه است. توزیع شرطی پارامتر b_i عبارت است از :

$$\prod(b_i / y_i, X, \beta_i) \propto L(\beta_i, b_i / \Sigma) * \Phi_j(b_i / \Sigma)$$

$$\prod \frac{\exp(-\lambda_{ij})\lambda_{ij}^{y_{ij}}}{\Gamma(y_{ij}+1)} \exp(-\frac{1}{2} b_i' \Sigma b_i)$$

$$\exp(\sum_{j=1}^{2}[y_{ij} b_i - \exp(x_{ij}')]) \exp(-\frac{1}{2} b_i' \Sigma b_i)$$

همان‌طور که مشاهده می‌شود، توزیع پسین به‌دست آمده برای پارامتر b_i دارای فرم بسته نیست، ازاین‌رو، برای نمونه‌گیری از چگالی شرطی b_i، می‌توان از الگوریتم متروپلیس- هستینگز استفاده کرد. به‌عنوان توزیع پیشنهاد از توزیع نرمال دومتغیری برای نمونه‌گیری به روش متروپلیس- هستینگز استفاده می‌شود.

٤-٢-توزیع شرطی کامل Σ^{-1}

توزیع شرطی پارامتر Σ^{-1} عبارت است از:

$$\prod f(\Sigma^{-1} / b_i) \propto f(\Sigma^{-1} / \nu_\Sigma, V_\Sigma) * \prod_{i=1}^{n} \Phi(b_i / \Sigma)$$

$$\prod f(\Sigma^{-1}/b_i) \propto f(\Sigma^{-1}/\nu_\Sigma, V_\Sigma) * \prod_{i=1}^{n} \Phi(b_i/\Sigma)$$

با توجه به توزیع پیشین ویشارت، توزیع شرطی کامل Σ نیز یک توزیع ویشارت با درجه آزادی $\nu_\Sigma + n$ است.

این توزیع، توزیع شناخته شده‌ای است و می‌توان مقدار Σ را در هر مرحله از یک توزیع ویشارت نمونه‌گیری کرد.

٤-٣- توزیع شرطی کامل β

توزیع شرطی پارامتر پسین β را می‌توان به صورت زیر تعریف کرد:

$$\prod(\beta_j|b_i,\Sigma,y,X) \propto \prod_{j=1}^{2}\prod_{i=1}^{n}(\beta_j|b_i,\Sigma,y_i,X)$$

با انجام محاسبات پیچیده توزیع شرطی پارامتر β را می‌توان برحسب تک‌تک مؤلفه‌های آن، یعنی βها تعریف کرد. در نتیجه، داریم:

$$\prod(\beta_j|b_j,\Sigma,y,X) \propto exp\left[-\frac{1}{2}\beta_j' V_\beta^{-1}\beta_j + \left(\sum_{i=1}^{n}\sum_{j=1}^{2} y_{ij}x_{ij}' + \beta_j' V_\beta^{-1}\right)\beta_j\right]$$

همان‌طور که مشاهده می‌شود، تابع به‌دست آمده دارای شکل مشخصی نیست، ازاین‌رو، ناچار به استفاده از الگوریتم متروپلیس- هستینگز هستیم. براساس الگوریتم متروپلیس- هستینگز، توزیع پیشنهادی برای بردار β_j یک توزیع t چندمتغیره با درجه آزادی ν_β است. برای انجام الگوریتم متروپلیس- هستینگز ابتدا لازم است یک مقدار پیشنهادی اولیه β_j^* را برای پارامتر β_j از توزیع $f_T(\beta_j|\widehat{\beta}_j, V, \upsilon_\beta)$ در نظر بگیریم. با در نظر گرفتن این مقدار پیشنهادی زنجیره از β_j به سمت مقدار پیشنهادی β_j^* با احتمال

$$\alpha = \min\left\{\frac{f_T(\beta_j \mid \hat{\beta}_j, V_\beta, \nu_\beta)\prod(\beta_j^* \mid y, X, b_j, \Sigma)}{f_T(\beta_j^* \mid \hat{\beta}_j, V_\beta, \nu_\beta)\prod(\beta_j \mid y, X, b_j, \Sigma)}, 1\right\}$$

حرکت خواهد کرد. مقدار پیشنهادی β_j^* با احتمال α پذیرفته می‌شود، در غیر این صورت، مقدار بعدی برای β_j^* در نظر گرفته و دوباره مراحل انجام می‌شود. در واقع، یک متغیر برنولی با احتمال α خواهیم داشت که براساس احتمال موفقیت آن می‌توانیم مقدار پیشتهادی β_j^* را انتخاب کنیم.

٤-٤- توزیع شرطی کامل β_β

در اینجا به توزیع شرطی پارامتر β_β نیاز است. برای این منظور با توجه به مطالب بیان شده، می‌توان گفت، ساختار توزیع دارای فرم بسته نیست، در نتیجه، در این حالت نیز به استفاده از الگوریتم متروپلیس– هستینگز نیاز است. برای نمونه‌گیری توزیع پیشنهادی t چندمتغیره است. در اینجا نیز مشابه قسمت‌های قبل یک مقدار پیشنهادی برای β_β، تحت عنوان متغیر β_β^* از تابع چگالی $f_T(\beta_j|\widehat{\beta}_J, V_\beta, \upsilon_\beta)$ در نظر گرفته می‌شود. در این حالت، زنجیر از مقدار β_β به β_β^* با احتمال زیر حرکت خواهد کرد.

$$\alpha = \min\left\{\frac{f_T(\beta_j \mid \hat{\beta}_j, V_\beta, \nu_\beta)\prod(\beta_j^* \mid \beta \text{و } \nu_\beta)}{f_T(\beta_j \mid \hat{\beta}_j^*, V_\beta, \nu_\beta)\prod(\beta_j \mid \beta, \nu_\beta)}, 1\right\}$$

در این حالت نیز با احتمال α می‌توانیم مقدار پیشنهادی β_β^* را انتخاب کنیم. در غیر این صورت، مقدار پیشنهادی β_β^* پذیرفته نمی‌شود.

٤-٥-توزیع شرطی کامل V_β^{-1}

توزیع شرطی بیزی ماتریس V_β^{-1} می‌تواند یک توزیع ویشارت با درجه آزادی $2 + \nu_\Sigma$ در نظرگرفته شود.

همچنین می‌توان به‌طور مستقیم با استفاده از نمونه‌گیری گیبس و گرفتن نمونه در هر مرحله از این توزیع متغیر آن را روز آمد. به‌طور مشابه توزیع شرطی بیزی V_Σ^{-1} نیز یک توزیع ویشارت با درجه آزادی $\nu_{0\Sigma} + \nu_\Sigma$ است.

بنابراین، با استفاده از نمونه‌گیری گیبس می‌توان این پارامتر را بهنگام کرد.

راه‌های متعددی برای معرفی این همبستگی‌ها در مدل پواسون دومتغیره وجود دارد. همان‌طور که بیان شد، در مطالعه‌های صورت گرفته گذشته ضریب همبستگی بین y1 و y2 (متغیر پاسخ اول و دوم) در مدل پواسون دومتغیره، محدود به اعداد مثبت بوده است. می‌توان نشان داد که با استفاده از این روش، علاوه بر کنترل بیش‌پراکندگی مدل، همبستگی بین دو متغیر y1 و y2 در مدل آمیخته حاصل تنها محدود به اعداد مثبت نیست و همبستگی‌های منفی را نیز پوشش می‌دهد. برای تحلیل داده‌های کاربردی از این روش استفاده می‌شود. در تمام مطالعات اشاره شده، سعی شده است خطای مدل پیشنهادی مورد ارزیابی و محاسبه قرار گیرد.

نظریه بیز قانونی است که برای ادغام اطلاعات مشاهده‌ای و اطلاعات حاصل از باورها، مجموعه‌ای واحد از اطلاعات بهنگام شده در رابطه با پارامتر یا آزمون فرضیه‌های مورد علاقه ایجاد می‌کند. در واقع، می‌توانیم استنباط دقیق‌تری را در مورد پارامتر مورد علاقه فراهم آوریم. باورهای پیشین درباره پارامترهای مجهول به‌طور عموم به‌عنوان توزیع‌های پیشین مطرح می‌شود. یکی از محدودیت‌های اصلی در کاربرد رهیافت‌های بیز، یافتن توزیع پسین بوده، زیرا محاسبه برخی از انتگرال‌ها حتی از حوزه بسیاری از روش‌های عددی پیشرفته نیز خارج است. در این گونه موارد ناچار به استفاده از روش‌های شبیه‌سازی برای برآورد توزیع پسین هستیم.

۵- معیارهای انتخاب بهترین مدل

پس از برآورد پارامترها و انجام تقریب‌های مناسب برای توزیع‌های پسین، به‌منظور بررسی درستی مدل پیشنهادی، در این بخش ملاک‌های اطلاع بیز (شوارتز) (BIC) و عامل بیز (BF) برای مقایسه بیزی مدل‌ها ارایه می‌شود. براساس تعریف عامل بیزی را با:

$$BF = \frac{P(y|M_0)}{P(y|M_1)}$$

نمایش می‌دهند، به‌طوری که در آن، $p(y|M)$ را تابع درست‌نمایی نوع دوم برای مدل M می‌نامند. افزایش یا کاهش مقدار تابع فوق می‌تواند در جهت تأیید یکی از مدل‌های ارایه شده حرکت کند.

یکی دیگر از ملاک‌هایی که می‌تواند برای بررسی و انتخاب یکی از مدل‌ها به کار رود، ملاک‌های اطلاع بیز (شوارتز) (BIC) بوده که براساس تعریف از رابطه زیر قابل محاسبه است:

$$BIC = D(\hat{\theta}) + Plog(n)$$

که در آن، $\hat{\theta}$ برآورد ماکزیمم درست‌نمایی θ، n تعداد مشاهدات در نمونه تصادفی و p تعداد پارامترهای قابل برآورد در مدل است. با استفاده از این معیار، مدلی که دارای کمترین مقدار معیار اطلاع بیز BIC باشد، به‌عنوان بهترین مدل انتخاب می‌شود. مزیت معیار اطلاع بیز BIC آن است که به توزیع پیشین بستگی ندارد.

در صورتی که ساختار مدل‌های پیشنهادی از یک پیچیدگی برخوردار باشد و بخواهیم از روش‌های محاسباتی پیچیده مانند روش مونت‌کارلوی زنجیره مارکوفی (MCMC) استفاده کنیم، می‌توانیم از ملاک پیچیده‌تری مانند ملاک اطلاع انحرافی DIC استفاده کنیم. براساس ملاک اطلاع انحرافی، مدلی که دارای کمترین مقدار DIC باشد، به‌عنوان بهترین مدل انتخاب می‌شود. این ملاک برای هر اندازه نمونه‌ای قابل استفاده بوده و به‌آسانی با روش‌های مختلف قابل اندازه‌گیری است.

٦- شبیه‌سازی و مدل‌سازی داده‌ها

در این بخش، برای بررسی درستی مدل‌های پیشنهادی، براساس داده‌های شبیه‌سازی مطالعه خود را انجام می‌دهیم و نتایج مربوط در ادامه می‌آید. داده‌ها که به‌طور شبیه‌سازی مربوط به یکی از بانک‌های کشور است، دارای دو متغیرتصادفی تعداد نوبت‌های استفاده از تسهیلات وام مشتریان برای ایجاد شغل و تعداد دفعات پرداخت جریمه دیرکرد وام توسط مشتری است که هر دو متغیر دارای توزیع پواسون و با یکدیگر همبسته هستند. در این ارتباط ابتدا باید اعلام کرد که برای ایجاد یک تحلیل مناسب به شیوه بیز و در حالت

داده‌های چندمتغیره شمارشی، به‌طور مشابه، اما با توزیع‌های پیشین مزدوج، اسکندری و نقی‌زاده (۱۳۸۴)، موضوع را مورد ارزیابی قرار دادند. نتایج تحقیق آنها که به شیوه بیز برای موضوع اشتغال انجام شد، بیان می‌کند که عوامل مختلفی، از جمله موقعیت جغرافیایی و جنسیت افراد می‌تواند بر تعداد شاغلان تأثیرگذار باشد. در این بررسی هرچند پاسخ‌ها رسته‌ای بود و مشاهده شد که عوامل ورودی بر ایجاد شغل در منطقه‌ای خاص تأثیرگذار است، اما در آن همبستگی بین پارامترهای متغیرهای رسته‌ای موجود، مورد بررسی قرار نگرفت. در این مطالعه با توجه به پیچیدگی محاسبات و به دلیل استفاده از توزیع‌های پیشین غیرمزدوج و نبود فرم بسته برای توزیع پسین، ابتدا برای داده‌ها، توزیع‌های پسین به صورت شبیه‌سازی و با استفاده از الگوریتم متروپلیس- هستینگز محاسبه شده‌اند، سپس، یک مدل آماری مناسب بیز برای بررسی همبستگی پنهان پارامترهای مربوط به دو متغیر پاسخ، ارایه شده و مورد بررسی قرار گرفته است. برای داده‌های شبیه‌سازی پیشنهادی که در آن، دو متغیر تعداد نوبت‌های استفاده از تسهیلات و تعداد دفعات پرداخت جریمه دیرکرد وام توسط مشتری در شعب بانک در تهران از طریق روش میدانی برای یک سال کاری جمع‌آوری شده است، مطالعه انجام گرفت. در این حالت، هرچند متغیرها دارای توزیع‌های پواسون مستقل از یکدیگر هستند، اما همبستگی پنهان معناداری بین پارامترهای آنها وجود دارد. برای بالا بردن دقت و با استفاده از مدل‌های خطی تعمیم‌یافته بیز که براساس الگوریتم متروپلیس- هستینگز حاصل شد، محاسبات آن انجام شد، برآورد پارامتر همبستگی را به شیوه بیز مورد بررسی قرار دادیم. در این رابطه، مقدار پارامتر α مربوط به الگوریتم متروپلیس- هستینگز برابر با ۶۰٪ شده است. در نتیجه، نمونه‌های شبیه‌سازی شده با احتمال ۶۰٪ تعیین شده و در محاسبات مورد استفاده قرار گرفته‌اند. در این مطالعه، اثر متغیرهای ورودی مانند میزان سرمایه اولیه، میزان تسهیلات پیشنهادی، نوع اشتغال‌زایی طرح به‌طور هم‌زمان بر تعداد نوبت‌های تسهیلات دریافتی و تعداد دفعات پرداخت جریمه دیرکرد وام مورد توجه قرار گرفته است. میزان تأثیر عوامل کمکی فوق بر متغیرهای پاسخ در جدول‌های شماره ۱ و ۲، آمده است.

جدول ۱- تأثیر عوامل کمکی بر تعداد نوبت‌های استفاده از تسهیلات (متغیر پاسخ اول)

ردیف	عوامل	کران پایین (۲/۵٪)	کران بالا (۹۷/۵٪)
۱	نوع اشتغال‌زایی	۰/۰۲۹۷۹	۰/۰۴۵۵
۲	میزان سرمایه اولیه	۰/۰۴۷۹	۰/۰۷۱۳
۳	میزان تسهیلات	۰/۰۲۲۳	۰/۱۸۱۳

جدول ۲- تأثیر عوامل کمکی بر تعداد دفعات پرداخت جریمه دیرکرد وام توسط مشتری (متغیر پاسخ دوم)

ردیف	عوامل	کران پایین (۲/۵٪)	کران بالا (۹۷/۵٪)
۱	نوع اشتغال‌زایی	۰/۰۲۵۶	۰/۳۴۲
۲	میزان سرمایه اولیه	۰/۰۱۶۶	۰/۴۴۶
۳	میزان تسهیلات	۰/۰۰۲۹۸	۰/۰۸۵۳

با توجه به منطبق بودن شرایط داده‌ها با مباحث نظری ارایه شده پس از به‌کارگیری مدل خطی تعمیم‌یافته بیز می‌توان مدل کاهش یافته پواسون- لگ‌نرمال را به داده‌ها برازاند. در تفسیر نتیجه‌های حاصل از داده‌های واقعی همان‌طور که اشاره شد، با مقایسه ملاک انتخاب بیز DIC مربوط به دو مدل پواسون- لگ‌نرمال (۴۲۰) با مدل پواسون–گامای دومتغیره (۵۰۱)، مدل پواسون - لگ‌نرمال مدل بهتری برای تفسیر نتایج در داده‌های واقعی مورد نظر است. برآورد پارامترهای ماتریس واریانس کوواریانس بیز مر بوط به پارامتر پنهان b_i را که دارای توزیع نرمال به صورت $b_i = (b_{i1}, b_{i2}) \sim N_2(0, \Sigma)$ است، برای دو متغیر پنهان b_{i1} و b_{i2} در جدول شماره ۳، ارایه می‌کنیم.

جدول ۳- برآورد پارامترهای ماتریس واریانس - کوواریانس مربوط به همبستگی پنهان

پارامتر	میانگین	انحراف معیار	کران پایین (۲/۵٪)	کران بالا (۹۷/۵٪)
b_{11}	۱۲۸/۱	۱۹۶/۴	۱۷/۲۱	۷۹۵/۲
b_{12}	۶۷/۳۹	۱۲۲/۹	۶۹	۱۳۴
b_{21}	۶۷/۳۹	۱۲۲/۹	۶۹	۱۳۴
b_{22}	۹۰/۶۱	۱۳۲/۷	۴/۸۹	۱۲۶/۳

نتایج تأیید همبستگی بیز پنهان، معناداری مثبتی را بین b_{i1} و b_{i2} نشان می‌دهد که این موضوع بیان می‌کند هر قدر تعداد دفعات دریافت وام بیشتر باشد، تعداد دفعات جریمه دیرکرد وام بیشتر خواهد بود. در این باره می‌توان نتایج زیر را ارایه کرد.

۷- نتیجه‌گیری

۱- میزان اشتغال‌زایی و توسعه در این مطالعه، تأثیر مستقیمی بر تعداد نوبت‌های دریافت تسهیلات داشته است، حال آنکه اشتغال‌زایی یکی از ارکان مهم بوده و این رابطه خود، مؤید این موضوع بوده که توزیع تسهیلات بین مشتریان از شاخص مهم در هرکاری، از جمله اشتغال‌زایی است.

۲- با توجه به ردیف دوم از جدول‌های شماره ۱ و ۲، میزان سرمایه اولیه که از طرف مشتری ارایه می‌شود، بر هر دو متغیر پاسخ تأثیر معناداری دارد. در واقع، می‌توان گفت، همبستگی پنهان پارامترهای دو متغیر پاسخ باعث تأثیرگذاری سرمایه اولیه بر دو متغیر پاسخ شده است.

۳- نوع اشتغال‌زایی (صنعتی،کشاورزی و خدمات) برای بانک (که توسط مشتری ارایه می‌شود)، بر پارامترهای هر دو متغیر پاسخ پواسون تأثیرگذار است.

۴- با توجه به گسترش کارآفرینی و شرکت‌های دانش‌بنیان و مانند آن که براساس ایده‌های افراد توسعه می‌یابد، لازم است یک مدیریت ساختاری برای ساماندهی این‌گونه رفتارها در جامعه ایجاد کرد. در این ارتباط مدل‌های آماری پیچیده‌تری به‌وجود خواهد آمد که لازم است توسط کارشناسان مختلف مورد استفاده قرار گیرد.

منابع

اسکندری، فرزاد و سیما نقی‌زاده (۱۳۸۴)، «تحلیل مدل‌های پویای بیز تعمیم‌یافته و مقایسه آن با روش‌های دیگر با کاربردی در بررسی اشتغال در کشور»، *پژوهشنامه اقتصادی ایران*، شماره ۳۴، دانشگاه علامه طباطبائی.

Aitchison, J., and Ho, C. H., (۲۰۱۰), The Multivariate Poisson-Lognormal Distribution, *Biometrika*, 4(4), pp.643-653.

Cowles, M. k., and Carlin, B. P., (1996), "Markov Chain Monte Carlo Convergence Diagnostics: .A Comparative Review", *Journal of the American Statistical Association*, 91, pp.883-904.

Johnson, N. L. and Kotz, S.(1969). *Distributions in Statistics: Discrete Distributions*, New York: John Wiley & Sons.

Karlis, D. Ntzoufras, I. (2005). "Bivariate Poisson and Diagonal Inflated". *Journal of Applied Statistics*, 30(1), pp.63-77.

Kazemnejad A., Zayeri F., Hamzah N.A., Gharaaghaji R., Salehi M.,A (2010). "Bayesian Analysis of Bivariate Ordered Categorical Responses using a Latent Variable Regression Model: Application to Diabetic Retinopathy Data", *Scientific Research and Essays*, Vol. 5, No. 11, pp.1273-1264.

Khafri, S. Kazemnejad, A. and Eskandari, F. (2008). "Hierarchical Bayesian Analysis of Bivariate Poisson Regression Model". *World Applied Sciences Journal* 4 (5) pp.667-675.

Kocherlakota, S, and Kocherlakota, K. (2001). "Regression in the Bivariate Poisson Distributions". *Communications in Statistics-Theory and Methods*, 30, pp.815-827.

Ntzoufras, J., Katsis, A., and Karlis, D., (2005). "Bayesian Assessment of the Distribution of the Insurance Claim Counts Using Reversible Jump MCMC". *North American Actuarial Journal*, 9(3), pp.90-108.

Ramooz, N. and Eskandari, F. (2016). "Analysis of Bivariate Correlated Data under the Poisson-Gamma Model". *Report and Opinion*. 8(6), pp.82-91.