تقریب توزیع صورت درجه دوم نامعین با رویکرد برآورد گشتاوری

قاسم رکابدار[۱]، رحیم چینی‌پرداز[۲] و بهزاد منصوری [۲]

[۱]گروه ریاضی، دانشگاه آزاد اسلامی، واحد آبادان

[۲]گروه آمار، دانشگاه شهید چمران اهواز

چکیده: در این مطالعه، از تابع چگالی نمایی چند پارامتری برای برآورد تابع چگالی صورت درجه دوم بردار متغیرهای نرمال استفاده شده است. به این منظور، صورت درجه دوم به صورت ترکیب وزنی از متغیرهای کای‌دو نامرکزی مستقل نشان داده شده و گشتاورهای آن از هر مرتبه‌ای محاسبه شده است. با استفاده از مقدار اشتاین در خانواده نمایی، پارامترهای تابع چگالی برآورد شده و با مثال‌های عددی نشان داده شده که این روش برای تقریب توزیع مناسب می‌باشد.

واژه‌های کلیدی: صورت درجه دوم، توزیع نرمال چندمتغیره، توزیع کای‌دو نامرکزی، آماره دوربین- واتسون.

۱ مقدمه

صورت درجه دوم متغیر نرمال یا مجموع وزنی از متغیرهای کای‌دو در حوزه‌های مختلف کاربردی آمار مورد استفاده قرار می‌گیرد. اگر صورت درجه دوم معین باشد، تمام وزن‌های متغیرهای کای‌دو مثبت و اگر نامعین باشد، برخی از وزن‌ها مثبت و برخی دیگر منفی هستند. از جمله کاربردهای صورت درجه دوم معین مثبت، می‌توان به توزیع آماره نیکویی برازش آندرسون-دارلینگ اشاره نمود که تحت فرضیه صفر به ترکیب وزنی از

°آدرس الکترونیک مسئول مقاله: قاسم رکابدار، gh.rekabdar@iauabadan.ac.ir

°کد موضوع‌بندی ریاضی (۲۰۱۰): ۱۱E۰۴, ۶۲H۳۰.

متغیرهای کای‌دو مرکزی همگرا می‌شود (گیلز، ۲۰۰۱). برای حالت نامعین می‌توان به توزیع متغیر ممیزی درجه دوم در حالت معلوم بودن پارامترها توجه نمود که در هر گروه این متغیر مجموع وزنی از متغیرهای کای‌دو نامرکزی است (کراکس و جوسنز، ۲۰۰۵، هوشمند، ۱۹۹۳).

برای تقریب توزیع صورت درجه دوم در توزیع نرمال یا ترکیب خطی از متغیرهای کای دو می‌توان به روش گرلاند (۱۹۵۵) با استفاده از تابع مشخصه اشاره نمود. با توجه به معلوم بودن گشتاورهای ترکیب خطی متغیرهای کای دو، استفاده از رویکرد تقریب تابع چگالی با گشتاورهای توزیع مورد توجه پژوهشگران بوده است. برای توزیع صورت درجه دوم از متغیر نرمال در حالت معین تقریب‌های پاتنایک (۱۹۴۹) و پیرسون (۱۹۵۹) به این دلیل که توزیع تقریبی، تابعی از توزیع کای‌دو مرکزی است؛ کاربرد فراوانی دارند. به ویژه روش پیرسون به این دلیل که علاوه بر میانگین و واریانس، چولگی توزیع را نیز در نظر می‌گیرد، برای محاسبه احتمال‌های تجمعی در دنباله بالایی مناسب‌تر است. در حالت نامعین هوشمند (۱۹۹۳) از روش پیرسون اصلاح شده برای برآورد خطای رده‌بندی تابع ممیزی درجه دوم استفاده کرده است. همچنین لیو و همکاران (۲۰۰۹) توزیع صورت درجه دوم نیمه معین مثبت را با توزیع کای‌دو نامرکزی تقریب زدند. در روش آن‌ها درجه آزادی و پارامتر نامرکزی متغیر کای‌دو توسط چهار کومولانت اول صورت درجه دوم برآورد شده است. در حالتی‌که صورت درجه دوم در توزیع نرمال چندمتغیره نامعین باشد، پروست و همکاران (۲۰۰۹) با نوشتن صورت درجه دوم نامعین به صورت تفاوت دو صورت درجه دوم مثبت مستقل بیان کرده و سپس توزیع تقریبی هر یک از صورت‌های مثبت را با استفاده از تقریب توزیع گاما چندجمله‌ای محاسبه کردند. در روش آن‌ها، تفاوت دو توزیع با استفاده از فرمول پیچش برای چگالی توأم صورت‌های مثبت درجه دوم محاسبه شده است. محسنی پور و پروست (۲۰۱۱) از تقریب توزیع صورت درجه دوم در حالتی که بردار نرمال دارای ماتریس کواریانس ویژه باشد با استفاده از چگالی‌های گاما و گاما تعمیم‌یافته استفاده کرده‌اند. بطور مشابه، ها و پروست (۲۰۱۳) به‌جای چندجمله‌ای گاما از تقریب چگالی هر یک صورت‌های درجه دوم با استفاده چندجمله‌ای لاگر[1] استفاده کرده‌اند. اخیراً رکابدار و چینی پرداز (۲۰۱۷) از روش تقریب بیشینه آنتروپی چگالی برای برآورد توزیع صورت درجه دوم استفاده کرده‌اند که در این روش تعداد گشتاورهای استفاده شده در تقریب معمولاً بیش از چهار گشتاور اول توزیع است.

در این مطالعه یک روش ساده برای تقریب تابع چگالی معرفی شده است که در آن بیش از چهار گشتاور برای تقریب بکار می‌رود. مزیت روش پیشنهادی این است که برآورد پارامترها در این روش ساده و سریع بوده و

[1] Laguerre Polynomial

می‌توان از تعداد دلخواه گشتاور یا کومولانت در برآورد توزیع صورت درجه دوم، استفاده کرد. ایده اصلی استفاده از مقدار اشتاین در خانواده نمایی است (هادسون، ۱۹۷۸). در این روش برای حالتی که گشتاورهای توزیع مشخص باشند، می‌توان چگالی تقریبی را با تابع چگالی نمایی چند پارامتری برآورد نمود.

در بخش ۲ صورت درجه دوم برای متغیرهای نرمال، به صورت ترکیب وزنی از متغیرهای کای‌دو نمایش داده شده است. در بخش ۳ برای برآورد تابع چگالی با استفاده از گشتاورهای صورت درجه دوم، روش پیشنهادی ارائه شده است. با ارائه دو مثال فرضی و تقریب توزیع آماره دوربین واتسون، در بخش ۴ عملکرد روش پیشنهادی بررسی شده است. در بخش ۵ بحث و نتیجه‌گیری آورده شده است.

۲ نمایش تصادفی صورت درجه دوم

نمایش تصادفی توابع پیچیده از متغیرهای تصادفی چند متغیره بر اساس متغیرهای تصادفی مستقل تک متغیره مرسوم نظیر کای‌دو، روش ساده‌تری را برای بررسی توزیع صورت‌های درجه دوم متغیرهای برداری فراهم می‌نماید. یک مزیت برای نمایش تصادفی صورت درجه دوم، امکان شبیه‌سازی مونت کارلو آن با تکرار زیاد است. همچنین می‌توان کومولانت‌ها را با استفاده از نمایش تصادفی از هر مرتبه‌ای محاسبه نمود.

فرض کنید $\boldsymbol{X} = (X_۰, \dots, X_d)$ بردار نرمال چندمتغیره با میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس Σ باشد. صورت درجه دوم $\boldsymbol{X}$ را به صورت $Q = X'AX$ تعریف می‌شود، که A یک ماتریس متقارن $d \times d$ است. صورت درجه دوم X را معین می‌نامند اگر همه مقادیر ویژه ماتریس A مثبت باشد. در غیر این‌صورت، اگر ماتریس A دارای مقادیر ویژه مثبت و منفی باشد صورت درجه دوم نامعین خواهد بود (ماسای و پروست، ۱۹۹۲). فرض کنید H ماتریس متعامدی باشد که ماتریس $\Sigma^{\frac{۱}{۲}} A \Sigma^{\frac{۱}{۲}}$ را قطری کند. با استفاده از قضیه تجزیه طیفی داریم

$$\Sigma^{\frac{۱}{۲}} A \Sigma^{\frac{۱}{۲}} = H \Lambda H', \tag{۱}$$

که در آن Λ ماتریس قطری با درایه‌های روی قطر $\lambda_۱ \geq \lambda_۲ \geq \cdots \geq \lambda_d$ است که به صورت نزولی مرتب شده‌اند. همچنین بردار ستونی

$$\boldsymbol{\delta} = H' \Sigma^{-\frac{۱}{۲}} \boldsymbol{\mu} \tag{۲}$$

را تعریف می‌کنیم. با تغییر متغیر $U = H'\Sigma^{-\frac{1}{2}}X$ آنگاه $\boldsymbol{U} \sim N_d(\boldsymbol{\delta}, I_d)$. در نتیجه نمایش تصادفی Q به صورت مجموع وزنی متغیرهای مستقل کای دو نامرکزی با یک درجه آزادی به صورت

$$Q = X'AX = U'H'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}HU = U'\Lambda U = \sum_{j=1}^{d}\lambda_j T_j, \qquad (۳)$$

است، که در آن $T_j \sim \chi_1^2(\delta_j^2)$. واضح است اگر $\boldsymbol{\mu} = 0$ آنگاه $\boldsymbol{\delta} = \mathbf{0}$ و صورت درجه دوم به صورت مجموع متغیرهای کای‌دو مرکزی مستقل هر کدام با یک درجه آزادی است. به دلیل مستقل بودن متغیرهای کای‌دو نامرکزی، با استفاده از کومولانت‌های توزیع کای‌دو نامرکزی (جانسون و همکاران، ۱۹۹۴)، به راحتی کومولانت‌های Q از طریق نمایش تصادفی آن به صورت :

$$\mathcal{K}_r = \sum_{j=1}^{d}\lambda_j^r\mathcal{K}_r(T_j) = 2^{r-1}(r-1)!\sum_{j=1}^{d}\lambda_j^r(1+r\delta_j^2), r = 1, 2, \ldots.$$

قابل محاسبه است. با استفاده از رابطه بازگشتی بین کومولانت‌ها و گشتاورهای توزیع می‌توان گشتاورهای Q را به صورت

$$\mu'_r = \sum_{j=0}^{r-1}\binom{r-1}{j}\mathcal{K}_r\mu'_j \qquad (۴)$$

محاسبه کرد (اسمیت، ۱۹۹۶).

۳ تقریب توزیع متغیر Q

از آنجا که صورت درجه دوم نامعین دارای وزن‌های منفی در مجموع وزنی است، چولگی متغیر Qمی‌تواند منفی و تقریب آن با توزیع‌هایی مناسب است که محدودیتی برای مقدار چولگی نداشته باشند. در این بخش، یک تابع نمایی انعطاف پذیر برای برآورد تابع چگالی احتمال Q در نظر گرفته می‌شود. تابع چگالی احتمال خانواده نمایی چند پارامتری به صورت

$$f_{\boldsymbol{\theta}}(x) = \exp\{\sum_{j=1}^{m}\eta_j(\boldsymbol{\theta})S_j(x) - B(\boldsymbol{\theta})\}h(x),$$

تعریف می‌شود. که η_j و B توابعی با مقادیر حقیقی از پارامترها بوده و S_j آماره‌های با مقادیر حقیقی و $\boldsymbol{x}$ نقطه‌ای در فضای نمونه χ از تکیه‌گاه تابع چگالی است. معمولاً برای سادگی η_j آن یک پارامتر فرض می‌شود

و تابع چگالی به صورت تابع نمایی متعارف

$$f_{\boldsymbol{\eta}}(x) = \exp\{\sum_{j=1}^{m} \eta_j S_j(x) - A(\boldsymbol{\eta})\}h(x), \tag{۵}$$

نوشته می‌شود. در محاسبات مربوط به گشتاورهای توزیع نمایی، هادسون (۱۹۷۸) برابری اشتاین را برای توزیع نمایی چند پارامتری توسعه داد. این رابطه در محاسبه گشتاورهای توزیع نمایی سودمند است. هادسون (۱۹۷۸) نشان داده است که در خانواده نمایی به صورت (۵)، اگر $g(x)$ هر تابع حقیقی مشتق‌پذیر باشد به طوری که $E|g(X)| < \infty$ آن‌گاه رابطه

$$E\{[\frac{h(X)}{h(X)} + \sum_{j=1}^{m} \eta_j S_j(X)]g(X)\} = -E(g(X)) \tag{۶}$$

برقرار است، که در آن $x \in (-\infty, \infty)$. اگر تکیه‌گاه X فاصله کران‌دار (a, b) باشد، آن‌گاه (۸) برقرار است اگر $\exp\{\sum_{j=1}^{n} \eta_j S_j(x)h(x)\} \to 0$ هنگامی که $x \to a$ یا $x \to b$ ، که $\to$ به معنای میل کردن یا نزدیک شدن است. به منظور استفاده از چگالی احتمال (۵)، برای تقریب تابع چگالی صورت درجه دوم، فرض کنید که برای $S_j(q) = q^j, \quad j = 1, \dots, m$ ، آن‌گاه تابع چگالی تقریبی صورت درجه دوم به صورت

$$f_{\boldsymbol{\eta},m}(q) = \exp\{\eta_1 q + \eta_2 q^2 + \cdots + \eta_m q^m - A(\boldsymbol{\eta})\}. \tag{۷}$$

خواهد بود. به عبارت دیگر در تابع نمایی، از تابع چندجمله‌ای از درجه m استفاده می‌شود. با قرار دادن $g(q) = q^{j+1}, j = 1, \dots, m$دستگاه معادلات خطی

$$\begin{cases} \eta_1 E(Q^2) + 2\eta_2 E(Q^3) + \cdots + m\eta_m E(Q^{m+1}) = -2E(Q) \\ \eta_1 E(Q^3) + 2\eta_2 E(Q^4) + \cdots + m\eta_m E(Q^{m+2}) = -2E(Q^2) \\ \vdots \\ \eta_1 E(Q^{m+1}) + 2\eta_2 E(Q^{m+2}) + \cdots + m\eta_m E(Q^{2m}) = -2E(Q^m) \end{cases}$$

با کمک برابری اشتاین تشکیل می‌شود. به عبارت دیگر گشتاور مرتبه k-ام را برابر با ترکیبی خطی از گشتاورهای $k+1$ تا $k+m$ با در نظر گرفتن ماتریس‌های

$$\boldsymbol{W} = \begin{bmatrix} E(Q^2) & 2E(Q^3) & \cdots & mE(Q^{m+1}) \\ E(Q^3) & 2E(Q^4) & \cdots & mE(Q^{m+2}) \\ \vdots & \vdots & \vdots & \vdots \\ E(Q^{m+1}) & 2E(Q^{m+2}) & \cdots & mE(Q^{2m}) \end{bmatrix}, \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} -2E(Q) \\ -3E(Q^2) \\ \vdots \\ -(m+1)E(Q^m) \end{bmatrix}$$

می‌توان دستگاه را به صورت ماتریسی $\boldsymbol{W\eta} = \boldsymbol{B}$ نوشت. چون $\boldsymbol{W} = \boldsymbol{DW}_0$ که در آن $D = diag\{1, 2, \ldots, m\}$ ماتریس قطری است و بنابراین $W_0\boldsymbol{\eta} = D^{-1}B$. نشان داده می‌شود ماتریس W_0 معین مثبت است. فرض کنید $\boldsymbol{a}$ برداری مخالف با صفر است آنگاه:

$$\boldsymbol{a'W}_0\boldsymbol{a} = [a_1, a_2, \ldots, a_m] \begin{bmatrix} E(Q^2) & 2E(Q^3) & \cdots & mE(Q^{m+1}) \\ E(Q^3) & 2E(Q^4) & \cdots & mE(Q^{m+2}) \\ \vdots & \vdots & \vdots & \vdots \\ (Q^{m+1}) & 2E(Q^{m+2}) & \cdots & mE(Q^{2m}) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

$$\begin{aligned}
&= a_1^2E(Q^2) + 2a_1a_2E(Q^3) + (a_2^2 + 2a_1a_3)E(Q^4) + \cdots a_m^2E(Q^{2m}) \\
&= \int_{-\infty}^{\infty} (a_1^2q^2 + 2a_1a_2q^3 + (a_2^2 + 2a_1a_3)q^4 + \cdots + a_n^2q^{2m})f(q)dq \\
&= \int_{-\infty}^{\infty} (a_1q + a_2q^2 + a_3q^3 + \cdots + a_mq^m)^2 f(q)dq \\
&= \int_{-\infty}^{\infty} p(q)f(q)dq
\end{aligned}$$

چون توزیع Q پیوسته است، بنابراین $E(p(Q)) > 0$. در نتیجه ماتریس W_0 معین مثبت است، بنابراین ماتریس W_0 همیشه وارون‌پذیر و دستگاه بالا همواره دارای جواب

$$\boldsymbol{\eta} = W^{-1}B \tag{8}$$

است. بعد از محاسبه $\boldsymbol{\eta}$ خواهیم داشت:

$$\begin{aligned}
\int_{-\infty}^{\infty} \exp\{\eta_1q + \eta_2q^2 + \cdots + \eta_mq^m - A(\boldsymbol{\eta})\}dq &= \\
\exp\{-A(\boldsymbol{\eta})\} \int_{-\infty}^{\infty} \exp\{\eta_1q + \eta_2q^2 + \cdots + \eta_mq^m\}dq &= 1
\end{aligned}$$

اگر از طرفین لگاریتم گرفته شود، رابطه

$$A(\boldsymbol{\eta}) = \ln(\int_{-\infty}^{\infty} \exp\{\eta_1 q + \eta_2 q^2 + \cdots + \eta_m q^m\} dq).$$

بدست می‌آید. در حالت کلی برای محاسبه تابع توزیع چگالی احتمال (۷) فرم بسته وجود ندارد و باید از انتگرال عددی برای محاسبه احتمال‌های تجمعی استفاده نمود. در عمل می‌توان، ابتدا نمایش تصادفی (۳) را با شبیه‌سازی مونت کارلو به اندازه زیاد، برای مثال 10^6 تکرار نموده و بازه‌های انتگرال‌گیری را مشخص نمود. برای مثال اگر a کمترین مقدار مشاهده شده در شبیه‌سازی باشد، در این صورت احتمال تجمعی در نقطه y به صورت

$$F(y) = \int_{a^*}^{y} f_{m,\boldsymbol{\eta}}(q) dq,$$

خواهد بود، که در آن a^* عددی کوچک‌تر از a است. برای تعیین درجه m از میان چگالی‌های موجود چگالی احتمالی را می‌توان انتخاب کرد که صدک‌های آن با صدک‌های شبیه‌سازی مونت کارلو کمترین فاصله را داشته باشند. به عبارت دیگر، صدک‌های مقادیر توزیع تجربی صورت درجه دوم از طریق تکرار به روش مونت کارلو محاسبه و t_α ها از رابطه

$$G(t_\alpha) = P(Q \leq t_\alpha) = \frac{\#(Q \leq t_\alpha)}{10^6},$$

محاسبه شدند که در آن $\#(.)$ به معنی تابع شمارشگر تعداد است. سپس صدک‌های توزیع برای m ثابت به صورت

$$F_m(t_\alpha) = \int_{-\infty}^{t_\alpha} f_{m,\boldsymbol{\eta}}(q) dq, \;\; m = 1, 2, \ldots,$$

محاسبه شد. پارامتر m به گونه ای انتخاب می‌شود که فاصله اقلیدسی

$$\Delta = \|F_m(t) - G(t)\| = \left\{ \sum_{j=1}^{99} (F_m(t_{0/01 \times j}) - G(t_{0/01 \times j}))^2 \right\}^{\frac{1}{2}}, \qquad (۹)$$

کمترین مقدار را داشته باشد. با وجود آن‌که محدودیتی برای انتخاب m وجود ندارد، در این مطالعه، برای انتخاب m، معادله (۹) برای $m \in \{1, \ldots, 30\}$ بررسی شده است.

۴ مثال‌های عددی

در این بخش، برای بررسی روش پیشنهادی در تقریب توزیع صورت درجه دوم، ابتدا یک مثال فرضی را در نظر گرفته شده که در آن توزیع دقیق صورت درجه دوم نامعین، مشخص باشد. ایمهوف (۱۹۶۱) در حالتی که ضرایب متغیرهای کای‌دو مرکزی زوج شده باشند و در هر زوج وزن‌ها برابر باشند توزیع دقیق را محاسبه کرده است. در این مثال، روش پیشنهادی با توزیع دقیق Q مقایسه خواهد شد. در مثال دوم یک بردار نرمال فرضی در نظر گرفته شده است و توزیع صورت درجه دوم آن در تکرار زیاد بررسی شده است. در مثال سوم برای داده‌های واقعی در یک مدل رگرسیونی، توزیع آماره دوربین- واتسون بررسی خواهد شد (پیوست الف). برنامه‌های رایانه‌ای مورد استفاده در این بخش با استفاده از نرم افزار متمتیکا نگاشته شده‌اند.

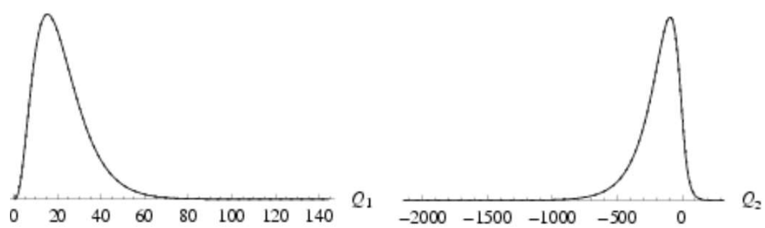
مثال ۱. فرض کنید در ترکیب وزنی از متغیرهای کای‌دو مرکزی وزن‌ها به صورت زوجی $\lambda_1 = \lambda_2 = 4, \lambda_3 = \lambda_4 = 3/5, \lambda_5 = \lambda_6 = 2/5, \lambda_7 = \lambda_8 = 1$ باشند. بنابراین نمایش تصادفی صورت درجه دوم به صورت

$$Q_1 = 4\chi^2_2 + 3/5\chi^2_2 + 2/5\chi^2_2 + \chi^2_2.$$

است. پارامترهای چولگی و کشیدگی Q_1 به ترتیب $\beta_1 = 1/17$ و $\beta_2 = -1/83$ هستند. برای صورت درجه دوم نامعین، فرض کنید وزن‌های متغیرهای کای‌دو مرکزی به صورت
$\lambda_1 = \lambda_2 = 16, \lambda_3 = \lambda_4 = 7, \lambda_5 = \lambda_6 = 3, \lambda_7 = \lambda_8 = 2, \lambda_9 = \lambda_{10} = 1,$
$\lambda_{11} = \lambda_{12} = -6, \lambda_{13} = \lambda_{14} = -14, \lambda_{15} = \lambda_{16} = -32, \lambda_{17} = \lambda_{18} = -70$ زوج شده باشند. آن‌گاه صورت درجه دوم دارای نمایش تصادفی

$$Q_2 = 16\chi^2_2 + 7\chi^2_2 + 3\chi^2_2 + 2\chi^2_2 + \chi^2_2 - 6\chi^2_2 - 14\chi^2_2 - 32\chi^2_2 - 70\chi^2_2$$

با شاخص‌های چولگی و کشیدگی به ترتیب $\beta_1 = -1/44$ و $\beta_2 = 0/60$ خواهد بود. کمترین مقدار Δ در معادله (۹) برای صورت درجه دوم معین Q_1 به ازای $m = 15$ به دست می‌آید که مقدار آن $0/0042$ است. همچنین برای صورت درجه دوم نامعین Q_2 کمترین مقدار $\Delta = 0/019$ بوده که به ازای ۱۷ است. لازم به توضیح است که در این مثال به دلیل مشخص بودن توزیع دقیق Q در معادله (۹)، به جای توزیع تجربی از توزیع دقیق استفاده شده است. در شکل (۱) منحنی تابع چگالی دقیق به همراه منحنی تابع چگالی برآورد



شکل ۱: تابع چگالی توزیع دقیق (منحنی توپر) و روش پیشنهادی (نقطه چین) .

شده به روش پیشنهادی ترسیم شده است. همان‌طور که از این صورت دیده می‌شود، استفاده از گشتاورهای زیاد و برآورد تابع چگالی نمایی به دست آمده به خوبی بر صورت دقیق منطبق شده‌اند، هر چند که هر دو صورت درجه دوم $Q_۱$ و $Q_۲$ چولگی شدیدی دارند.

مثال ۲. فرض کنید برای صورت درجه دوم $Q_۴ = X'AX$ ، ماتریس صورت درجه دوم به‌صورت

$$A = \begin{bmatrix} ۱ & -۱ & ۰ & ۰ \\ -۱ & ۲ & ۱ & ۰ \\ ۰ & ۱ & ۲ & -۲ \\ ۰ & ۰ & -۲ & -۱۴ \end{bmatrix},$$

است. همچنین بردار Xدارای توزیع نرمال ۴ متغیره با بردار میانگین $\boldsymbol{\mu}' = (۵, ۴, -۴, -۱۲)$ و ماتریس کواریانس معین

$$\Sigma = \begin{bmatrix} ۵ & ۰/۵ & ۰/۵ & ۰/۵ \\ ۰/۵ & ۸ & ۰/۵ & ۰/۵ \\ ۰/۵ & ۰/۵ & ۴ & ۰/۵ \\ ۰/۵ & ۰/۵ & ۰/۵ & ۱۵ \end{bmatrix},$$

باشد. بنابرین ماتریس قطری مقادیر ویژه ماتریس $\Sigma^{\frac{۱}{۲}} A \Sigma^{\frac{۱}{۲}}$ به‌صورت

$$\Lambda = diag\{۲۰/۸۰۴۱, ۷/۸۸۵۹۳, ۱/۱۹۶۲۱, -۲۱۲/۸۸۶\},$$

و صورت درجه دوم Qدارای نمایش تصادفی نامعین

$$\begin{aligned} Q &\sim ۲۰/۸۰۴۱\chi_۱^۲(۰/۰۱۵۲۸۷) + ۷/۸۸۵۹۳\chi_۱^۲(۰/۱۶۷۷۸۵) \\ &+ ۱/۱۹۶۲۱\chi_۱^۲(۱۱/۶۰۴۳) - ۲۱۲/۸۸۶\chi_۱^۲(۱۰/۳۶۴۸) \end{aligned}$$

است. پارامترهای چولگی و کشیدگی Q به‌ترتیب $\beta_۱ = -۰/۸۹۵$ و $\beta_۲ = ۱/۹۲$ هستند. برای مشخص کردن درجه m در روش پیشنهادی با مشخصه اشتاین در خانواده نمایی، کمترین مقدار فاصله اقلیدسی برای $m = ۲۹$ درجه به‌دست آمده است. در جدول ۱ احتمال‌های تجمعی روش‌های تقریبی باهم مقایسه شده‌اند. به

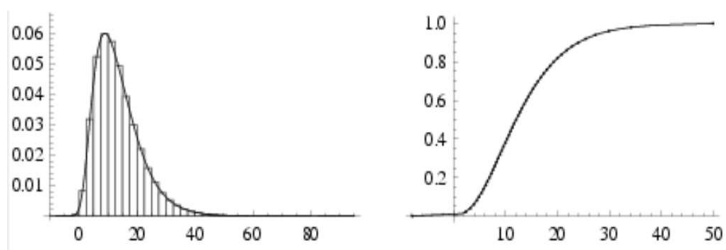
جدول ۱: احتمال های تجمعی Q : شبیه‌سازی (SIM)، پیرسن اصلاح شده (P) و روش پیشنهادی (St)

St	P	t	SIM	St	P	t	SIM
۰/۵۹۹۹	۰/۶۰۴۶۳	−۱۸۲۳/۸۶	۰/۶۰	۰/۰۰۹۹	۰/۰۱۰۲	−۶۵۰۱/۳۹	۰/۰۱
۰/۶۹۹۴	۰/۷۰۵۸۴	−۱۴۲۶/۹۵	۰/۷۰	۰/۰۴۹۹	۰/۰۴۹۵	−۴۹۹۲/۲۴	۰/۰۵
۰/۷۹۹۴	۰/۸۰۵۸۴	−۱۱۵۳/۲۹	۰/۸۰	۰/۰۰۹۹	۰/۰۹۸۶	−۴۲۶۸/۷۰	۰/۱۰
۰/۸۹۹۵	۰/۹۰۲۵۹	−۱۷۴۹/۶۲	۰/۹۰	۰/۱۹۹۹	۰/۱۹۹	−۳۴۶۲/۹۶	۰/۲۰
۰/۹۴۹۱	۰/۹۴۸۹۸	−۷۴۹/۵۹۹	۰/۹۵	۰/۳۰۰۱	۰/۲۹۹۲	−۲۹۳۶/۲۱	۰/۳۰
۰/۹۸۷۳	۰/۹۸۵۲۳	−۱۲۲/۸۹۷	۰/۹۹	۰/۵۰۰۱	۰/۵۰۲۶	−۲۱۸۵/۴۲	۰/۵۰

این منظور با شبیه‌سازی مونت کارلو در $۱۰^۶$ تکرار، چندک‌های t محاسبه شده‌اند و در این مقادیر احتمال‌های تجمعی روش پیشنهادی با روش پیرسون اصلاح شده به روش هوشمند (۱۹۹۳) مقایسه شده است. همان‌طور که از این جدول مشاهده می‌شود، روش پیشنهادی دارای احتمال‌های تجمعی نزدیک‌تری به صدک‌های شبیه‌سازی مونت کارلو است.

مثال ۳. یکی از فرضیه‌های اساسی در مدل‌های رگرسیونی خطی عدم وجود خودهمبستگی در میان خطاهای مدل است. وجود خودهمبستگی در میان خطا مدل بیانگر اطلاعات بیشتری در داده‌ها است که توسط پژوهشگر در نظر گرفته نشده است. توزیع آماره دوربین واتسون را می‌توان با استفاده از صورت درجه دوم نامعین تقریب زد (اسمال، ۱۹۹۳). برای بررسی عملکرد روش پیشنهادی در تقریب آماره دوربین–واتسون برای سال‌های ۱۳۷۶ لغایت ۱۳۸۹، آمارهای نرخ بیکاری به عنوان متغیر وابسته (y) و متغیرهای مستقل نرخ سود بانکی $(x_۱)$ ، درآمد نفتی$(x_۲)$ و شاخص تولیدات صنعتی $(x_۳)$ در نظر گرفته شده است. بنابراین، $T = ۱۴$

و $T - d = ۱۰$. مقدار آماره دوربین واتسون $\hat{D} = ۱/۰۰۲$ و مقدار خودهمبستگی نمونه‌ای مرتبه اول بین خطاهای مدل رگرسیون خطی چندگانه $\hat{\rho} = ۰/۴۷$ بوده است. مقادیر ویژه در مجموع وزنی متغیرهای کای‌دو در معادله (الف۔۲) به ترتیب ۰/۵۴−، ۰/۰۶−، ۰/۲۴، ۰/۸۲، ۱/۳۵، ۱/۷۶، ۲/۰۷، ۲/۴۳، ۲/۵۶، ۲/۸۸، محاسبه شده‌اند. برای روش پیشنهادی مقدار $m = ۲۹$ در نظر گرفته شده است، که مقدار



شکل ۲: نمودار چگالی و احتمال‌های تجمعی آماره $\hat{D}$.

Δ در (۹) برابر ۰/۰۲۶ است. در شکل ۲ نمودارهای تابع چگالی احتمال برآورد شده و تابع احتمال تجمعی تقریبی برای روش پیشنهادی به همراه نمودار بافت‌نگار و نمودار نقطه ای احتمال‌های تجمعی شبیه‌سازی مجموع وزنی متغیرهای کای‌دو به روش مونت کارلو با $۱۰^۶$ تکرار نمایش داده شده است. نمودار چگالی تقریبی بر روی بافت‌نگار داده‌ها شبیه‌سازی شده کاملاً منطبق شده است و عدم تقارن را به خوبی نشان داده است. احتمال‌های تجمعی تجربی نیز بر نمودار احتمال تجمعی تقریب زده شده قرار گرفته‌اند. مقادیر احتمال برای فرض $H_۰ : D = ۲$ در برابر $H_۱ : D < ۲$ در جدول ۲ گزارش شده‌اند. با روش پیشنهادی در سطح خطای یک درصد پذیرفته می‌شود که بین خطاهای مدل خودهمبستگی مثبت وجود دارد. با تقریب نرمال (گجراتی، ۲۰۰۳) نیز وجود خودهمبستگی مثبت در سطح خطای ۵ درصد معنی‌دار است.

جدول ۲: تقریب مقدار احتمال برای $\hat{D}$ با روش پیشنهادی

روش پیشنهادی	آماره	مقدار $-p$
نرمال	۱/۸۶	۰/۰۳۰۹
شبیه‌سازی	—	۰/۰۰۱۴
پیشنهادی	—	۰/۰۰۱۹

بحث و نتیجه‌گیری

در این مقاله یک روش جدید برای برآورد تابع چگالی معرفی شده است. روش پیشنهاد شده برای تقریب تابع چگالی صورت درجه دوم نامعین، روشی ساده و دقیق بوده که در آن می‌توان به هر تعداد دلخواه از گشتاورهای توزیع استفاده نمود. چون برآورد پارامترها در این روش نیازی به الگوریتم‌های عددی ندارد و با معادله (۸) همواره قابل محاسبه است، به‌دست آوردن تقریب‌ها سریع می‌باشند. معمولاً درجه مورد نیاز در این روش حداقل برابر ۱۵ است و می‌تواند به عنوان نقطه ابتدایی برای به‌دست آوردن چگالی پیشنهاد شود. این روش می‌تواند در تقریب چگالی متغیرهایی که گشتاورهای معلوم داشته اما تابع چگالی آن‌ها نامشخص است، به کار گرفته شود.

تقدیر و تشکر

نویسندگان مقاله از داوران محترم مقاله، سردبیر و ویراستار مجله که با راهنمایی و توصیه‌های مفید خود بهبود مقاله را هموار کردند، قدردانی می‌نمایند.

پیوست الف: توزیع تقریبی آماره دوربین-واتسون

مدل‌های رگرسیون خطی چندگانه به صورت

$$y_t = \beta_۰ + \beta_۱ x_{t,۱} + \cdots + \beta_d x_{t,d-۱} + \varepsilon_t, \quad t = ۱, \ldots, T,$$

یا به صورت ماتریسی $y = X\beta + \varepsilon$ نوشته می‌شود. برای این مدل، آماره دوربین واتسون بر اساس وجود خودهمبستگی از مرتبه اول در میان جمله خطا است. یعنی فرض می‌شود رابطه:

$$\varepsilon_t = \rho\varepsilon_{t-۱} + \nu_t, \quad |\rho| < ۱,$$

برقرار است، که در آن ρ ضریب همبستگی بین خطاهای متوالی ε_t و ε_t و ν_t دارای توزیع نرمال با میانگین صفر و واریانس ثابت $\sigma^۲$ است. برآوردگر ρ ، خود هبستگی از مرتبه اول و به صورت

$$\hat{D} = \frac{eA^*e}{e'e} \qquad (۱۰)$$

است. بردار e باقیمانده‌های کمترین توان‌های دوم معمولی است که با معادله $e = My$، به دست می‌آید. ماتریس $M = I_T - X(X'X)^{-1}X'$ ماتریس متقارن و خودتوان با رتبه $T-d$ ، ماتریس A^* در معادله (۱۰) دارای عناصر $a_{11} = a_{dd} = 1$، $a_{ii} = 2$ برای $i = 2, \dots, d-1$، $a_{ij} = -1$ اگر $|i-j| = 1$ و $a_{ij} = 0$ اگر $|i-j| \geq 2$ است (اسمال، ۱۹۹۳). فرض کنید P ماتریسی $T \times (T-d)$ باشد که با متعامدسازی $(T-d)$ ستون مستقل خطی m_j از ماتریس M داده شود، بنابراین $M = PP'$. در نتیجه $P'X = o$ و $P'P = I_{T-d}$. چون $MY \sim N_n(o, \sigma^2 I_T)$ و $z = \sigma^{-1}MY \sim N_n(o, I_T)$ است بنابراین می‌توان تابع توزیع احتمال Dرا در نقطه d به صورت

$$
\begin{aligned}
P(\hat{D} \leq d) &= P(\frac{Y'M'A^*MY}{Y'MY} \leq d) \\
&= P((\frac{(\beta'X' + \varepsilon)PP'A^*PP'(X\beta + \varepsilon)}{(\beta'X' + \varepsilon')PP'(X\beta + \varepsilon)} \leq d) \\
&= P(\frac{z'P'A^*Pz}{z'z} \leq d) = P(z'(P'A^*P - dI_{T-d})z \leq 0).
\end{aligned}
$$

نوشت با استفاده از مطالب بخش (۲)، اگر $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{T-d}$ مقادیر ویژه ماتریس $P'A^*P - dT_{T-d}$ باشند، آنگاه

$$P(\hat{D} \leq d) = P(\sum_{j=1}^{T-d} \lambda_j \chi_1^2 \leq 0).$$

رابطه تقریبی آماره دوربین- واتسون با $\hat{\rho}$یعنی خودهمبستگی مرتبه اول نمونه ای، به صورت $\hat{D} \approx 2(1-\hat{\rho})$ است، چون $\hat{\rho}$ به صورت تقریبی دارای توزیع نرمال $N(0, T)$ است، بنابراین متغیر $\sqrt{T}(1 - \hat{D}/2)$ به صورت تقریبی دارای توزیع نرمال استاندارد است (گجراتی، ۲۰۰۳).

مراجع

[1] Croux, C. and Joossens, K., (2005), Influence of Observations on the Misclassification Probability in Quadratic Discriminant Analysis, *Journal of Multivariate Analysis,* **96**, 384-403.

[2] Giles, D. E. A., (2001), A Saddle Point Approximation to the Distribution Function

of the Anderson-Darling Test Statistic, *Communications in Statistics -Simulation and Computation* **30**, 899-905.

[3] Gujarati, D., (2003), *Basic Econometrics*, McGraw Hill, New York.

[4] Gurland, J., (1955), Distribution of Definite and of Indefinite Quadratic Forms, *The Annals of Mathematical Statistics,* **26**, 122-127.

[5] Ha, H. T. and Provost, S. B., (2013), AN Accurate Approximation to the Distribution of a Linear Combination of Non-Central Chi-Square Random Variables, *REVSTAT*, **11**, 231-254.

[6] Houshmand, A. A., (1993), Misclassification Probabilities for Quadratic Discriminant Function, *Communications in Statistics-Simulation and Computation*, **22**, 81-98.

[7] Hudson, H. M., (1978), A Natural Identity for Exponential Families with Application in a Multi-Parameter Estimation, *Annals of Statistics*, **6**, 473-484.

[8] Imhof, J. P., (1961), Computing the Distribution of Quadratic Forms in Normal Variables, *Biometrika,* **48**, 419-426.

[9] Johnson, N. L., Kotz, S. and Balakrishnan, N., (1994), *Continuous Univariate Distributions,* Vol. 1. New Yourk: Wiley Series in Probability and Mathematical Statistics, Wiley.

[10] Liu, H., Tang, Y. and Zhang, H.H., (2009), A New Chi-square Approximation to the Distribution of Non-Negative Definite Quadratic Forms in Non-Central Normal Variables, *Computational Statistics and Data Analysis*, **53**, 853-856.

[11] Mathai, A. M. and Provost, S. B., (1992), *Quadratic Forms in Random Variables, Theory and Applications,* New York, Marcel Dekker.

[12] Mohsenipour, A. A. and Provost, S., B., (2011), On approximating the Distribution of Indefinite Quadratic Expressions in Singular Normal Vectors, *Acta Et Commentationes Universitatis Tartuensis De Mathematica*, **15**, 61-86.

[13] Patnaik, P. B., (1949), The Non-Central and F-Distributions and Their Applications, *Biometrika,* **36**, 202-232.

[14] Pearson, E. S., (1959), Note on An Approximation to the Distribution of Non-Central , *Biometrika,* **46**, 364.

[15] Provost, S. B., Ha, H. T. and Sanjel, D., (2009), On Approximating the Distribution of Indefinite Quadratic Forms, Statistics: *A Journal of Theoretical and Applied Statistics*, **43**, 597-609.

[16] Rekabdar, G., and Chinipardaz, R., (2017), Approximating the Distribution of Indefinite Quadratic Forms in Normal Variables by Maximum Entropy Density Estimation, *Journal of Modern Applied Statistical Methods*, **16**, 359-377.

[17] Small, J. P., (1993), The Limiting Power of Point Optimal Autocorrelation Tests, *Communications in Statistics-Theory and Methods*, **22**, 3907-3916.

[18] Smith, P. J., (1995), A Recursive Formulation of the Old Problem of Obtaining Moments From Cumulants and Vice Versa, *American Statistician*, **49**, 217-219.

On Approximating the Distribution of Indefinite Quadratic Forms by Moments Approach

Rekabdar[1], G., Chinipardaz[2], R., Mansouri[2], B.

[1]Department of Mathematics, Islamic Azad University, Abadan Branch, Abadan, Iran.

[2]Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz , Iran.

Abstract:

In this study, the multi-parameter exponential family of distribution has been used to approximate the distribution of indefinite quadratic forms in normal random vectors. Moments of quadratic forms can be obtained in any orders in terms of representation of the quadratic forms as weighted sum of non-central chi-square random variables. By Stein's identity in exponential family, we estimated parameters of probability density function. The method handled in some examples and we indicated this method suitable for approximating the quadratic form distribution.

Keywords: Quadratic form, Multivariate Normal Distribution, Non-Central Chi-Square Distribution, Durbin-Watson Statistic.

Mathematics Subject Classification (2010): 62H30, 62H10.