توزیع‌های دوجمله‌ای منفی و پوآسون پارامتر- آماسیده

فرشته مؤمنی[۱] محمدحسین علامت‌ساز[۲]

چکیده

در عمل ضرورت استفاده از خانواده‌های غنی‌تر توزیع‌های احتمال گسسته کلاسیک دوجمله‌ای منفی و پوآسون احساس می‌شود. در مقاله حاضر با وارد کردن پارامتر اضافی ρ، نشان خواهیم داد که توزیع‌های جدید، که آنها را اصطلاحاً پارامتر آماسیده می‌نامیم، برای مدل‌بندی داده‌های شمارشی همبسته که فراپراکنش (واریانس بزرگتر از میانگین) دارند مناسب‌ترند. کاربرد آنها را در داده‌های واقعی خسارت بیمه نشان داده و خصوصیات این توزیع‌ها را نیز با توزیع‌های متناظر کلاسیک مقایسه می‌کنیم.

واژه‌های کلیدی : توزیع برنولی، توزیع دوجمله‌ای منفی، توزیع پوآسون، توزیع صفر آماسیده، توزیع‌های پارامتر آماسیده

۱. مقدمه

از توزیع‌های مهمی که در شمارش کاربرد وسیعی دارند می‌توان به توزیع‌های هندسی، دوجمله‌ای منفی و پوآسون اشاره کرد. اخیراً ضرورت تعمیم این توزیع‌ها با توجه به مناسب نبودن توزیع‌های کلاسیک در مدل‌بندی برخی داده‌های فراوانی یا داده‌های وابسته در زمینه‌های مختلف اقتصادسنجی، دارایی، بیمه، زیست‌شناسی و غیره توسط باورز و همکاران [۱] ، رولسکی و همکاران[۵] وین کلن[۶] و کولو و همکاران[۳] مورد تأکید قرار گرفته است. بویژه، توزیع‌های کلاسیک در مدل‌بندی داده‌های بیمه مورد تردید قرار گرفته‌اند.

در اینجا ضمن معرفی توزیع‌های پارامتر آماسیده هندسی، دوجمله‌ای منفی و پوآسون، توزیع‌های احتمال و توابع مولد آنها را به دست آورده و با استفاده از یک مجموعه داده واقعی مربوط به تعداد خسارات بیمه حوادث در استان مازندران ارجحیت این توزیع‌ها را بر توزیع‌های کلاسیک متناظر نشان خواهیم داد.

۲. توزیع های صفر- آماسیده

مطالعه ما برتوزیع های گسسته اصلاح شده درصفر(صفر- آماسیده) پایه‌گذاری شده است، که برای مدل سازی شمارش هایی که به طورنامتناسبی شامل فراوانی بزرگی (یا کوچکی) درصفرهستند مورداستفاده قرارمی گیرد. فرض کنید X یک متغیرتصادفی دلخواه با مقادیر صحیح نامنفی دلخواه باشد، بطوریکه

$$P(X=j)=p_j \quad , \quad j=0,1,2,\dots \quad , \quad \sum_{j=0}^{\infty} p_j=1$$

و $P_X(t)=E(t^X)$ تابع مولد احتمال آن باشد . حال اگر در این توزیع صفرها را با نسبت اضافی $\rho\in(0,1)$ در نظر بگیریم نسبتهای باقی مانده به طورمناسبی کاهش می یابد.

به این ترتیب متغیر جدید Y از X حاصل می‌شود که آن را اصطلاحاً صفر – آماسیده X گوییم. به عبارت دیگر، داریم

$$P(Y=0)=\rho+(1-\rho)p_0$$
$$P(Y=j)=(1-\rho)p_j \quad , \quad j=1,2,\dots \qquad (1)$$

پس تابع مولد احتمال متغیر صفر آماسیده Y به صورت زیر است :

$$P_Y(t)=\rho+(1-\rho)P_X(t) \qquad (2)$$

توزیع‌های صفر - آماسیده جانشین‌های مناسبی برای مدل سازی نمونه های خوشه ای که در آن جامعه شامل دو زیرجامعه، یکی فقط حاوی صفر و دیگری مشاهده مقادیرتوزیع گسسته ای است، بکار می‌رود. این امکان وجود دارد که ρ مقداری کمتر از صفر بگیرد (کاهش نسبت صفر) مشروط برآن که $P(Y=0)\geq 0$ باشد. در این صورت باید $\rho>\frac{-p_0}{1-p_0}$ و در نتیجه با توجه به اینکه $|\rho|\leq 1$ مورد نظر است لذا باید $\max\left\{-1,\frac{-p^0}{1-p_0}\right\}\leq\rho\leq 0$.

۱- دانشجوی کارشناسی ارشد دانشگاه اصفهان

۲- عضو هیأت علمی گروه آمار دانشگاه اصفهان

اگر $\rho = 1$، آنگاه توزیع صفر- آماسیده متناظر، توزیع تباهیده درصفر می باشد و اگر $\rho = 0$ باشد آنگاه هیچ تغییری در(۲) ایجاد نمی‌شود، یعنی، $P_Y(t) = P_X(t)$.

حال با استفاده از توزیع های صفر- آماسیده توزیع‌های پارامتر- آماسیده را تعریف می کنیم .

۳. توزیع هندسی پارامتر- آماسیده

فرض کنید $\{W_1, W_2, \ldots\}$ دنباله ای از متغیرهای مستقل و همتوزیع برنولی با احتمال موفقیت π باشند. یعنی،

$$W_k = \begin{cases} 0, & \text{با احتمال } (1-\pi) \\ 1, & \text{با احتمال } \pi \end{cases}$$

برای $k = 1, 2, \ldots$ و $\pi \in (0, 1)$ باشد .

حال دنباله صفر- آماسیده متناظر با آن $\{\overline{W}_1, \overline{W}_2, \ldots\}$ را در نظر بگیرید که برای $k = 1, 2, \ldots$ بنابر(۲) به صورت زیر تعریف می‌شود :

$$\overline{W}_k = \begin{cases} 0, & \text{با احتمال } (1-\pi)(1-\rho)+\rho \\ 1, & \text{با احتمال } (1-\rho)\pi \end{cases}$$

که درآن $\max\left\{-1, -\frac{1-\pi}{\pi}\right\} \le \rho < 1$. فرض کنید متغیرتصادفی V نمایشگر تعداد آزمایش‌هایی باشد که برای رسیدن به اولین پیروزی در دنباله متغیرهای برنولی مستقل جدید $\{\overline{W}_1, \overline{W}_2, \ldots\}$ مشاهده می شود. در این صورت تابع جرم احتمال متغیر تصادفی V به صورت زیر است :

$$P(V=k) = [(1-\pi)(1-\rho)+\rho]^{k-1}(1-\rho)\pi \quad , k=1,2,.. \qquad (۳)$$

متغیر تصادفی V دارای توزیع هندسی معمولی بر روی اعداد صحیح مثبت، $V \sim Ge(\pi^*)$، است که در آن

$$\pi^* = (1-\pi)(1-\rho)+\rho = 1-(1-\rho)\pi$$

اکنون متغیر تصادفی X را با استفاده از توزیع V به صورت زیر تعریف می کنیم :

$$p_0 = P(X=0) = \pi \qquad (۴)$$

$$p_k = P(X=k) = (1-\pi)[(1-\pi)(1-\rho)+\rho]^{k-1}(1-\rho)\pi, \; k=1,2,..$$

واضح است که تابع مولد احتمال X به صورت زیر است.

$$P_X(t) = \sum_{K=0}^{\infty} t^k \; P(X=k)$$

$$= \pi + [p\pi(1-\pi)(1-\rho)t]\sum_{k=1}^{\infty}\{[(1-\pi)(1-\rho)+\rho]t\}^{k-1}$$

$$= \frac{\pi(1-t\rho)}{1-t[1-\pi+\pi\rho]} \qquad (۵)$$

تعریف۱. به متغیرتصادفی X تعریف شده در(۴) (یا (۵)) توزیع هندسی پارامتر- آماسیده با پارامترهای

$$\rho \in \left(\max\left\{-1, -\frac{1-\pi}{\pi}\right\}, 1\right) \text{ و } \pi \in (0,1)$$

می گوییم و آن را با نماد $X \sim IGe(\pi, \rho)$ نشان می‌دهیم.

به سهولت دیده می‌شود که میانگین و واریانس توزیع $IGe(\pi, \rho)$ برابرند با :

$$E(X) = \frac{1-\pi}{\pi(1-\rho)},$$

$$Var(X) = \frac{(1-\pi)(1+\pi\rho)}{\pi^2(1-\rho)^2}$$

تذکر ۱ : در صورتی که $\rho = 0$ در نظر گرفته شود، به سادگی از (۴) دیده می‌شود که $\{p_k\}$ همان توزیع هندسی کلاسیک تعداد شکست ها تا اولین پیروزی می‌باشد و لذا تابع مولد احتمال (۵)، $E(Y)$ و $V(Y)$ در بالا به همان مقادیر کلاسیک هندسی تبدیل می‌شوند.

تذکر ۲ : توزیع هندسی کلاسیک از توزیع‌های تک - مدی گسسته محسوب می‌شود. یادآور می‌شویم که توزیع گسسته $\{p_k\}$ را در نقطه $n = n_0$ تک مدی گویند هرگاه

$$\begin{cases} p_n \ge p_{n-1} & , \quad n \le n_0 \\ p_{n+1} \le p_n & , \quad n \ge n_0 \end{cases}$$

لذا چون توزیع هندسی $\{p_k\}$، تابعی نزولی از k است، یک توزیع تک مدی در مبدا خواهد بود. حال این سئوال مطرح می‌شود که آیا توزیع هندسی پارامتر آماسیده (۴) نیز این خاصیت را حفظ می‌کند. پاسخ مثبت است. این موضوع را به سادگی می‌توان ثابت کرد.

چون

$$\Delta p_k = p_{k+1} - p_k = \begin{cases} \pi[(1-\pi)(1-\rho)-1] \ , & k=0 \\ -\pi(1-\pi)(1-\rho)^2[(1-\pi)(1-\rho)+\rho]^k \ , & k=1,2,\ldots \end{cases}$$

در نتیجه دنباله $\{\Delta p_k\}$ حداکثر در یک نقطه $(k=1)$ می‌تواند (از + به −) تغییر علامت می‌دهد. بنابراین، توزیع تعمیم یافته نیز تک مدی خواهد بود.

تذکر ۳ : مشهور است که توزیع هندسی کلاسیک تنها توزیع گسسته‌ای است که دارای خاصیت فقدان حافظه است. یعنی رابطه زیر را برقرار می‌کند

(۶)

$$\forall x \geq 0 \ \& \ x_0 \geq 0: \quad P(X \geq x + x_0 \mid X \geq x_0) = P(X \geq x)$$

لذا بدیهی است که توزیع هندسی پارامتر آماسیده $\mathrm{IGe}(\pi, \rho)$ نمی‌تواند دارای خاصیت فقدان حافظه باشد، مگر البته برای $\rho = 0$ که به همان توزیع هندسی کلاسیک تغییر می‌یابد. لیکن می‌توان نشان داد که به‌طور مشابه در اینجا رابطه زیر که آن را "خاصیت فقدان حافظه نوع ρ" می‌خوانیم برقرار است.

$$\forall x \geq 0 \ \& \ x_0 \geq 0: \quad P(X \geq x + x_0 \mid X \geq x_0) = (1-\rho)P(X \geq x) + \rho P(X \geq x \mid X > 0) \quad (۷)$$

توجه کنید که در حالت $\rho = 0$، رابطه (۷) به همان رابطه (۶) تغییر می‌یابد.

٤. توزیع دوجمله ای منفی پارامتر- آماسیده

فرض کنید r یک عدد صحیح مثبت و $X_1, X_2, \ldots, X_r$ متغیرهای تصادفی مستقل و هم توزیع $\mathrm{IGe}(\pi, \rho)$ مشخص شده توسط (۴) باشند. آنگاه، تعریف می‌کنیم:

تعریف۲. متغیرتصادفی $Y = X_1 + X_2 + \cdots + X_r$ دارای توزیع دوجمله ای منفی پارامتر- آماسیده با پارامترهای $r \geq 1$، $\rho \in \left(\max\left\{-1, \frac{-1-\pi}{\pi}\right\}, 1\right)$، $\pi \in (0,1)$ است وآن را با نماد $Y \sim \mathrm{INB}(\pi, \rho, r)$ نشان می‌دهیم.

با توجه به تابع مولد احتمال متغیرهای تصادفی مستقل وهم توزیع $X_1, X_2, \ldots, X_r$ در (۵)، تابع مولد احتمال متغیر تصادفی $Y \sim \mathrm{INB}(\pi, \rho, r)$ به صورت زیر به دست می‌آید :

$$P_Y(t) = \left[\frac{\pi(1-t\rho)}{1-t(1-\pi+\pi\rho)}\right]^r \qquad (۸)$$

توزیع احتمالات دوجمله‌ای منفی پارامتر آماسیده صورت پیچیده‌ای دارد که در قضیه زیر به دست می‌آید[۴].

قضیه ۲. تابع جرم احتمال متغیرتصادفی $Y \sim INB(\pi, \rho, r)$ عبارتست از:

$$P(Y=y) = \pi^r \sum_{y_1, y_2, \ldots} \binom{y_1+y_2+\cdots+r-1}{y_1, y_2, \ldots, r-1} [(1-\pi)(1-\rho)]^{y_1+y_2+\cdots} \rho^{y_1+2y_2+\cdots} \qquad (۹)$$

که در آن $y = 0, 1, \ldots$ و جمع بندی روی اعداد صحیح نامنفی $y_1, y_2, \ldots$ است به طوری که، $y_1 + 2y_2 + \cdots = y$

اثبات . با استفاده از معادلات ترکیباتی، از تابع مولد احتمال(۸) نهایتا بدست می آوریم :

$$P_Y(t) = \left[\frac{\pi(1-t\rho)}{1-t(1-\pi+\pi\rho)}\right]^r = \frac{\pi^r}{\left[1-(1-\pi)(1-\rho)t\frac{1}{1-t\rho}\right]^r}$$

$$= \pi^r \sum_{m=0}^{\infty} \binom{-r}{m} \left[-(1-\pi)(1-\rho)t\left(1+t\rho+t^2\rho^2+\cdots\right)\right]^m$$

$$= \pi^r \sum_{m=0}^{\infty} \binom{m+r-1}{m} \left[(1-\pi)(1-\rho)t\right]^m \sum_{m_1, m_2, \ldots} \binom{m}{m_1, m_2, \ldots} (t\rho)^{m_2+2m_3+\cdots}$$

در مجموع آخر، جمعبندی روی اعداد صحیح نامنفی $m_1, m_2, \ldots$ انجام می‌گیرد به طوری که، $m_1 + m_2 + \cdots = m$. حال با استفاده از رابطه

$$\binom{m+r-1}{m}\binom{m}{m_1, m_2, \ldots} = \binom{m+r-1}{m_1, m_2, \ldots, r-1}$$

بدست می‌آوریم:

$$P_Y(t)=\pi^r \sum_{m=0}^{\infty}\left[(1-\pi)(1-\rho)t\right]^m$$

$$\sum_{m_1, m_2, \ldots}\binom{m+r-1}{m_1, m_2, \ldots, r-1}(t\rho)^{m_2+2m_3+\ldots}$$

با جایگذاری $m_i = y_i \ , i \geq 1$ و $m = y - \sum_{j=1}^{\infty}(j-1)y_j$

در عبارت اخیر سرانجام به دست می‌آوریم:

$$P_Y(t) = \sum_{y=0}^{\infty} t^y \ P(Y=y)$$

که درآن $P(Y=y)$ تابع جرم احتمال توزیع دوجمله‌ای منفی پارامتر- آماسیده تعیین شده در (۹) می‌باشد.

تذکر۴: میانگین و واریانس توزیع $INB(\pi, \rho, r)$ عبارتند از :

$$E(Y)=\frac{r(1-\pi)}{\pi(1-\rho)} \quad , \quad Var(Y)=\frac{r(1-\pi)(1+\pi\rho)}{\pi^2(1-\rho)^2}$$

که در حالت $\rho = 0$ به همان میانگین و واریانس توزیع دوجمله‌ای منفی کلاسیک $NB(r,\pi)$ تبدیل می‌شود.

۵. توزیع پوآسون پارامتر- آماسیده

در این جا با استفاده از یافتن حدود عبارات (۸) و (۹) به طوری که ،

$$r \to \infty \ \ \& \ \ \pi \to 1 : r(1-\pi)=\lambda \qquad (10)$$

تابع جرم احتمال و تابع مولد احتمال یک توزیع جدید را به دست می آوریم . در اینجا $\lambda > 0$ مقدار ثابتی است. گرفتن حد تابع مولد با استفاده از قضیه زیر تعیین می شود .

قضیه۳. تحت شرایط حدی (۱۰) رابطه زیر برقرار است .

$$\lim_{r\to\infty}\lim_{\pi\to 1} P_Y(t) = \exp\left[\frac{\lambda(t-1)}{1-\rho t}\right] \qquad (11)$$

که در آن $P_Y(t)$ تابع مولد احتمال (۸) می باشد.

اثبات . با گرفتن لگاریتم از دو طرف (۸) داریم :

$$\ln P_Y(t) = r\{\ln[1-(1-\pi+\pi\rho t)]-\ln[1-t(1-\pi+\pi\rho)]\}$$

با به کارگیری بسط تیلور تابع لگاریتمی $\ln(1-x)$ ، بعد از چند تبدیل ساده بدست می‌آید :

$$\ln P_Y(t)=r(1-\pi)(t-1)\Big\{1+\frac{1}{2}\left[2\pi\rho t+(1-\pi)(t+1)\right]$$

$$+\frac{1}{3}\left[3(\pi\rho t)^2\right.$$

$$\left.+3\pi\rho t(t+1)(1-\pi)+(t^2+t+1)(1-\pi)^2\right]+\cdots\Big\}$$

اکنون با استفاده از شرایط حدی (۱۰) خواهیم داشت :

$$\lim_{r\to\infty}\lim_{\pi\to 1}\ln P_Y(t)=\lambda(t-1)[1+\rho t+(\rho t)^2+\ldots]$$

$$=\frac{\lambda(t-1)}{1-\rho t}$$

بنابراین :

$$\lim_{r\to\infty}\lim_{\pi\to 1} P_Y(t) = \exp\left[\frac{\lambda(t-1)}{1-\rho t}\right]$$

حال با توجه به قضیه بالا ، متغیر تصادفی متناظر را به صورت زیر تعریف می کنیم .

تعریف ۳. گوییم متغیرتصادفی Z دارای توزیع پوآسون پارامتر- آماسیده با پارامترهای $\lambda > 0$ و $\rho \in [0,1]$ است، و آن را با نماد $Z \sim IPo(\lambda, \rho)$ نشان می‌دهیم، اگر تابع مولد احتمال آن بتواند به صورت زیر نمایش داده شود :

$$P_Z(t) = \exp\left[\frac{\lambda(t-1)}{1-\rho t}\right] \qquad (12)$$

مشابه با حالت دوجمله ای منفی ، تابع جرم احتمال توزیع پوآسون پارامتر- آماسیده را توسط قضیه زیر بدست می‌آوریم .

قضیه ۴. تابع جرم احتمال متغیرتصادفی Z با توزیع $IPo(\lambda, \rho)$ عبارتست از:

$$P(Z=z)=\sum_{z_1,z_2,\ldots}\frac{e^{-\lambda}}{z_1!z_2!\cdots}\left[\lambda(1-\rho)\right]^{z_1+z_2+\cdots}\rho^{z_2+2z_3+\cdots}\quad,\ z=0,1,\ldots \tag{13}$$

که در آن جمع‌بندی روی اعداد صحیح نامنفی z_1 , $z_2,\ldots$ انجام می‌شود، به طوری که $z_1+2z_2+3z_3+\cdots=z$.

اثبات . از تابع مولد احتمال (۱۲) داریم :

$$P_Z(t)=\exp\left[\lambda(t-1)(1+\rho t+\rho^2t^2+\ldots)\right]$$
$$=\exp\left\{\lambda[-1+(1-\rho)t+\rho(1-\rho)t^2+\rho^2(1-\rho)t^3+\ldots]\right\}$$

پس با استفاده از بسط تیلور تابع نمایی e^x داریم :

$$P_Z(t)=\sum_{n=0}^{\infty}\frac{\lambda^n\left[-1+(1-\rho)t+\rho(1-\rho)t^2+\rho^2(1-\rho)t^3+\cdots\right]^n}{n!}$$
$$=\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\sum_{n_0,n_1,\ldots}\binom{n}{n_0,n_1,\ldots}(-1)^{n_0}(1-\rho)^{n_1+n_2+\cdots}$$
$$\times\rho^{n_2+2n_3+\cdots}t^{n_1+2n_2+3n_3+\cdots}$$

که جمع‌بندی روی اعداد صحیح نامنفی n_0 , $n_1,\ldots$ انجام می‌شود به طوری که، $n_0+n_1+\cdots=n$. با جایگذاری $n=z-\sum_{j=0}^{\infty}(j-1)z_j$, $n_i=z_i$, $i\geq 0$ در عبارت اخیر می توان نوشت :

$$P_Z(t)=\sum_{z=0}^{\infty}t^z\sum_{z_1,z_2,\ldots}\frac{\left[\lambda(1-\rho)\right]^{z_1+z_2+\cdots}}{z_1!z_2!\cdots}\rho^{z_2+2z_3+\cdots}\sum_{z_0=0}^{\infty}\frac{(-\lambda)^{z_0}}{z_0!}$$

که در آن $z=0,1,2,\ldots$ می باشد . بنابراین، چون $\sum_{z_0=0}^{\infty}\frac{(-\lambda)^{z_0}}{z_0!}=e^{-\lambda}$ داریم :

$$P_Z(t)=\sum_{z=0}^{\infty}t^z\ P(Z=z)$$

که در آن $P(Z=z)$ برای $z\geq 0$ توسط (۱۳) داده شده است .

تذکر۵. میانگین و واریانس توزیع $IPo(\lambda,\rho)$ عبارتند از :

$$E(Z)=\frac{\lambda}{1-\rho}\qquad,\qquad Var(Z)=\frac{\lambda(1+\rho)}{(1-\rho)^2}$$

تذکر۶. ملاحظه می کنیم که توزیع $IPo(\lambda,\rho)$ منطبق با توزیع پولیا - ایپلی می باشد. (به[۲] مراجعه شود).

۶. کاربرد

در این بخش توزیع‌های پوآسون کلاسیک و پارامتر آماسیده را به یک گروه داده‌های واقعی که مربوط به تعداد ادعاهای خسارت بیمه حوادث در سال ۱۳۸۱ است و براساس ۱۶۷۶۰ سند (تعداد بیمه‌شدگان) در استان مازندران تهیه شده است برازش داده، نشان می‌دهیم که توزیع پوآسون پارامتر آماسیده قویاً مناسب‌تر است. جدول فراوانی داده‌ها به صورت زیر است :

k	۰	۱	۲	۳	۴≤
o_k	۱۳۷۷۲	۲۶۳۱	۳۱۸	۳۴	۵

و لذا داریم $\overline{X}=2/02$ و $S^2=0/21502$.

۶. ۱. برازش توزیع پوآسون کلاسیک : $Po(\lambda)$

فرض کنید π_k احتمال متناظر رده k ام و e_k فراوانی مورد انتظار رده kام باشند. در این صورت با برآورد پارامتر $\hat{\lambda}=\overline{X}=2/02$، آماره نیکویی برازش χ^2 برای توزیع پوآسون در جدول (۱) در زیر محاسبه می‌شود:

k	o_k	π_k	e_k	$(e_k-o_k)^2/e_k$
۰	۱۳۷۷۲	۰/۸۱۶۹۳	۱۳۶۹۱/۷۷	۰/۴۷
۱	۲۶۳۱	۰/۱۶۵۱۸	۲۷۶۸/۴۷	۶/۸۲
۲	۳۱۸	۰/۰۱۶۷۱	۲۷۹/۹۹	۵/۱۶
۳	۳۴	۰/۰۰۱۱۲	۱۸/۸۶	۱۲/۱۵
≥۴	۵	$6/02\times10^{-5}$	۱/۰۱	۱۵/۷۶
جمع	۱۶۷۶۰	۱	۱۶۷۶۰	$\chi^2=40/4$

جدول ۱: حالت پوآسون کلاسیک

پس چون مقدار مشاهده‌ای $\chi^2=40/4$ بزرگتر از نقطه بحرانی $\chi^2_{0/05,3}=7/815$ است فرض پوآسون بودن توزیع داده‌ها در سطح ۵

درصد، حتی در سطح معنی‌دار ۰/۵ درصد ($\chi^2_{0.005,2}$=۱۲/۸۳۸)، قویاً رد می‌شود.

۶. ۲. برازش توزیع پوآسون پارامتر- آماسیده :

$IPo(\lambda, \rho)$ با توجه به میانگین و واریانس توزیع $IPo(\lambda, \rho)$ که در تذکر۵ در بخش ۵ داده شده است، برآوردهای گشتاوری پارامترهای λ و ρ در این حالت عبارتنداز :

$$\hat{\rho} = \frac{S^2 - \overline{X}}{S^2 + \overline{X}} \quad \& \quad \hat{\lambda} = \frac{2\overline{X}^2}{S^2 + \overline{X}}$$

لذا، در اینجا داریم $\hat{\lambda}$ = ۰/۱۹۵۹ و $\hat{\rho}$=۰/۰۳۱ و بنابراین جدول محاسبه مقدار آماره نیکویی برازش χ^2 در این حالت به صورت جدول زیر به دست خواهد آمد.

لذا برخلاف حالت قبل در اینجا چون χ^2=۱/۴۰۷ و $\chi^2_{0.05,2}$=۵/۹۹۱، فرض صفر مبنی بر برازیدن توزیع پوآسون پارامتر آماسیده به داده‌ها قویاً مورد پذیرش قرار می‌گیرد.

خاطر نشان می‌کنیم که در اینجا مثبت بودن برآورد $\hat{\rho}$ = ۰/۰۳۱ نشان می‌دهد که مشاهده صفرها بیشتر از آن است که توسط توزیع پوآسون کلاسیک پیش‌بینی می‌شود.

k	o_k	π_k	e_k	$(e_k - o_k)^2/e_k$
۰	۱۳۷۷۲	۰/۸۲۲	۱۳۷۷۷/۲	۰/۰۰۲
۱	۲۶۳۱	۰/۱۵۶۱	۲۶۱۶/۳۵	۰/۰۸۰
۲	۳۱۸	۰/۰۱۹۶	۳۲۹/۵۳	۰/۴۰۳
۳	۳۴	۰/۰۰۲	۳۳/۶۴	۰/۰۰۴
≥۴	۵	$1/95\times10^{-4}$	۳/۲۶۸	۰/۹۱۸
جمع	۱۶۷۶۰	۱	۱۶۷۶۰	χ^2 =۱/۴۰۷

جدول۲: حالت پوآسون پارامتر - آماسیده

مراجع

[۱] بردبار عشرت‌آبادی، محمد ،۱۳۸۲، فنون محدودسازی افشای داده‌های آماری و کاربرد آنها، پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.

[2] Cox, L. H., 1980, Suppression methodology and statistical disclosure Control, *Journal of the American Statistical Association*, 75, 337-385.

[3] Cox L. H. 1987, A constructive procedure for unbiased controlled rounding, *Journal of the American Statistical Association*, 82, 520-524.

[4] Fienberg, S. E., Markov, U. G., and Steel, R. J. 1998, Disclosure limitation using perturbation and related methods for categorical data, *Journal of Official Statistics*, 14, 485-502.

[5] Gouweleeuw, J. M., Kooiman, P., Willenborg, L. C. J., and De Wolf, P. P., 1998, Post randomisation for statistical disclosure control: Theory and implementation, *Journal of Official Statistics*, 14, 463-474.

[6] Kooiman, P., Willenborg, L. C. J. and Gouweleeuw, J. M., 1997, PRAM: A Method for Disclosure Limitation of Microdata, *Statistics Netherlands,* The Netherlands, Research Paper No. 97-5.

[7] Willenborg, L. C. J. and De Waal, T., 1998, *Statistical Disclosure Control in Practice*, Springer-Verlag, New York.