مجله‌ی بررسی‌های آمار رسمی ایران
سال ۲۱، شماره‌ی ۱، بهار و تابستان ۱۳۸۹، صص ۱۳ – ۲۹

تحلیل بیزی توزیع درامد خانوار

فرزاد اسکندری[†,*] و رقیه قبادی[‡]

[†] دانشگاه علامه‌طباطبایی
[‡] وزارت بازرگانی

چکیده: مطالعه‌ی وضعیت براورد پارامترهای توزیع پارتو توسط آمارشناسان مختلفی مورد بررسی قرار گرفته است. اما این بررسی‌ها بیش‌تر به شیوه‌ی فراوانی‌گرا[۱] انجام پذیرفته است. هدف این مقاله، انتخاب توزیع‌های پیشین مناسب، برای پارامترهای توزیع پارتو و تعیین براورد پارامترهای آن به روش بیز می‌باشد. این کار را وقتی پارامترهای مربوط به توزیع‌های پیشین معلوم باشد، انجام خواهیم داد. در این وضعیت علاوه بر این‌که براورد پارامترها با استفاده از توزیع پسین به دست آمده است، با محاسبه‌ی واریانس‌های براوردگرها، مقایسه‌ای بین روش‌ها، تحت پیشین مزدوج انجام داده و آن را در یک مسئله‌ی کاربردی نیز مورد بررسی قرار داده‌ایم.

واژگان کلیدی: مقادیر فرین؛ پارامترهای اولیه؛ مدل‌بندی بیزی؛ توزیع پارتو؛ توزیع پسین.

۱– مقدمه

یکی از خانواده‌های مهم از توزیع‌ها که در آمار دارای اهمیت زیادی است. خانواده‌ی توزیع‌های مقادیر فرین[۲] می‌باشد. این توزیع‌ها دارای تکیه‌گاه ثابتی نیستند و فضای متغیر تصادفی و پارامتر مجهول جامعه به یکدیگر وابسته می‌باشند. از آنجایی که همه‌ی اطلاعات موجود در مشاهده با همه‌ی اطلاعاتی که در آماره‌ی بسنده[۳] وجود دارد هم‌ارز است، لذا اگر از این جامعه نمونه‌ی تصادفی n تایی اختیار گردد، در آن صورت همه‌ی اطلاعاتی که در مشاهده وجود دارد هم‌ارز اطلاعات موجود در کم‌ترین مقدار نمونه و یا بیش‌ترین مقدار نمونه می‌باشد. یکی از توزیع‌های مهم که عضو این خانواده است و در عمل نیز بسیار کاربرد دارد توزیع پارتو است. بررسی این توزیع از سال‌های دور مورد توجه

دریافت: ۱۳۸۸/۵/۴ ، پذیرش: ۱۳۸۹/۲/۲۳
[*] نویسنده‌ی عهده‌دار مکاتبات

آمارشناسان و همچنین اقتصاددانان بوده است، به‌طوری که می‌توان به کارهایی که گالامبوس [۳]، گندنکو [۴] و جکینسون [۵] انجام داده‌اند اشاره نمود. اما اولین فردی که به‌طور رسمی این توزیع را معرفی کرد ویلفردو پارتو (۱۹۶۴)، اقتصاددان مشهور ایتالیایی بوده است که این توزیع را به‌عنوان توزیع درامد بیان نمود [۹]. همچنین چمپرنون ثابت کرد که توزیع درامد، صرف نظر از توزیع اصلی آن می‌تواند با توزیع پارتو تقریب شود [۲]. در خصوص مطالعه‌ی توزیع درامد و ارتباط آن با توزیع آماری پارتو، افراد زیادی مطالعه نمودند که از آن جمله می‌توان به پرسکه [۱۰]، لەوی و سولومون [۶] اشاره نمود. قابل‌توجه است که این مطالعات اکثراً به شیوه‌ی فراوانی‌گرا[۴] انجام گرفته است. در حالت چندمتغیره نیز ماردینا مطالعات خود را بر روی این توزیع انجام داده و خواص مطلوبی را برای براورد پارامترها بدست آورده است [۸]. همچنین برای بررسی براوردگرهای ماکسیمم درستنمایی پارامترهای بهین توزیع، می‌توان به مقاله وینگو اشاره کرد [۱۱]. تحلیل بیزی براورد پارامترهای توزیع پارتو در دو حالت که پارامترهای اولیه‌ی آن معلوم و یا مجهول باشد کم‌تر از ۳ دهه است که انجام گرفته است. از کارهایی که با نگرش بیز بر روی توزیع پارتو انجام گرفته است می‌توان به مقاله‌ی مالیک در سال ۱۹۷۰ به‌عنوان اولین مقاله اشاره نمود که توزیع پارتو در حالتی که پارامتر مقیاس آن معلوم می‌باشد مورد بررسی قرار گرفته است [۷]. در حالتی که فرض وابستگی بین توزیع‌های پیشین پارامترهای توزیع پارتو وجود داشته باشد؛ آرنو و پرس مطالعه وسیعی را انجام دادند [۱]. نکته‌ای که در تمام این مطالب وجود داشته است تعیین شکل توزیع‌های پیشین است که به حالت معمول و با فرض معلوم بودن پارامترهای اولیه انجام گرفته است. در این مقاله با به کارگیری توزیع‌های پیشین مزدوج[5] برای پارامترهای مدل، به تحلیل بیزی توزیع پارتو پرداخته، آن را در یک مثال واقعی آن را به کار برده و نتایج حاصل را به‌وسیله‌ی نسبت واریانس‌ها مقایسه می‌کنیم. این مقاله شامل ۴ بخش می‌باشد. در بخش دوم این مقاله به تحلیل بیزی توزیع درامد پارتو وقتی پارامتر مقیاس، δ، معلوم باشد پرداخته می‌شود. در بخش ۳ تحلیل را وقتی هر دو پارامتر مجهول می‌باشد انجام خواهیم داد. سرانجام در بخش ۴ به کاربرد مدل‌ها می‌پردازیم.

۲- تحلیل بیزی توزیع پارتو وقتی پارامترهای اولیه‌ی توزیع پیشین و δ معلوم می‌باشد.

فرض کنید X_1 ، ...، X_n یک نمونه‌ی تصادفی از تابع توزیع پارتو به صورت زیر باشند:

$$F(x) = P(X \le x) = 1 - \left(\frac{\delta}{x}\right)^{\alpha}.$$

به‌سادگی می‌توان مشخص نمود که چگالی آن برابر است با

$$f(x) = \frac{\alpha\delta^{\alpha}}{x^{\alpha+1}}, \qquad x \ge \delta > 0, \quad \alpha > 0. \tag{1}$$

همچنین با اندکی محاسبه می‌توان نشان داد براوردگرهای ماکسیمم درست‌نمایی α و δ عبارت‌اند از:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} LnX_i - nLn\hat{\delta}}, \qquad \hat{\delta} = X_{(1)}.$$

این براوردگرها وقتی هر دو پارامتر نامعلوم باشند بدست آمده‌اند. می‌توان ثابت نمود توزیع این دو براوردگر مستقل از یکدیگر است.

پس از تعیین براوردهای ماکسیمم درست‌نمایی پارامترها می‌توانیم با استفاده از بازه‌های اطمینان مناسب یک براورد بازه‌ای برای هر پارامتر بدست آوریم. یکی از روش‌های مهم برای تعیین بازه‌ی اطمینان روش کمیت محوری می‌باشد، و بعد از انجام محاسباتی ساده می‌توان گفت توزیع براوردگری که به‌روش کمیت محوری حاصل می‌شود، خی‌دو با $(2n-2)$ درجه آزادی می‌باشد که در (۲) آمده است.

$$2\alpha\left[\sum_{i=1}^{n} LnX_i - nLnX_{(1)}\right] \sim \chi^2(2n-2) \tag{2}$$

وقتی دو پارامتر مجهول باشد با کمک رابطه‌ی (۲)، داریم:

$$\mathrm{P}\left[X^{۲}_{\frac{r}{۲}}(۲n-۲)<۲\alpha\left(\sum_{i=۱}^{n}\mathrm{Ln}X_{i}-n\mathrm{Ln}X_{(۱)}\right)<X^{۲}_{۱-\frac{r}{۲}}(۲n-۲)\right]=۱-r \quad ۰\leq r\leq ۱$$

و در نتیجه بازه‌ی اطمینان برای پارامتر α عبارت است از

$$\left(\frac{\chi^{۲}_{\frac{r}{۲}}(۲n-۲)}{۲\left(\sum_{i=۱}^{n}LnX_{i}-nLnX_{(۱)}\right)},\frac{\chi^{۲}_{۱-\frac{r}{۲}}(۲n-۲)}{۲\left(\sum_{i=۱}^{n}LnX_{i}-nLnX_{(۱)}\right)}\right) \qquad (۳)$$

از این بازه‌ی اطمینان می‌توان برای پذیرش یا عدم پذیرش مقدار α در یک وضعیت مشخص استفاده نمود.

اکنون زمانی که می‌خواهیم با استفاده از نگرش بیز پارامترها و استنباط‌های دیگر را در مورد پارامترهای توزیع پارتو براورد کنیم، حالت‌های مختلفی بوجود می‌آید به‌طوری که پیچیدگی کار در این بخش کمی بیش‌تر از حالت فراوانی‌گرای آن می‌باشد. همان‌طور که می‌دانیم تحلیل بیزی بر اساس توابع زیان مختلف می‌تواند نتایج متفاوتی را به همراه داشته باشد، لذا در ادامه به انتخاب چند توزیع پیشین و چند تابع زیان مختلف می‌پردازیم. آمارشناسان تحقیقات بسیاری انجام دادند تا مشخص شود که به راستی کدام توزیع پیشین نسبت به توزیع پیشین دیگر بهتر است؟ در این باره راه‌حل‌های موضعی نیز بدست آمد. اما می‌توان گفت هنوز هیچ آمارشناسی نتوانسته است به این پرسش پاسخ روشنی دهد. پس با ذکر توضیحات قبل سعی می‌کنیم تا در مورد تحلیل بیزی توزیع پارتو نیز این نکات مورد توجه قرار بگیرد.

اکنون توزیع پارتو در فرمول (۱) را در نظر بگیرید. از آنجایی که فقط پارامتر α مجهول و δ معلوم است پس فقط یک توزیع پیشین برای α در نظر گرفته می‌شود.

در حالتی که δ عددی معلوم باشد پیشنهاد می‌شود توزیع نمایی یا گامای دم‌برده به‌عنوان توزیع پیشین در نظر گرفته شود. با توجه به این که تحت هر مقدار δ استنباط آماری یکسان است، پس تحت $\delta=۱$ تابع درست‌نمایی عبارت است از

$$L(\alpha)=\alpha^{n}\exp\left\{-(\alpha+۱)\sum_{i=۱}^{n}LnX_{i}\right\}$$

اگر توزیع پیشین نمایی دوپارامتری اختیار شود، خواهیم داشت

$$\pi(\alpha)=\lambda e^{-\lambda(\alpha-\mu)} \underset{(\mu,\infty)}{I}(\alpha), \qquad I(\alpha)=\begin{cases}1 & \mu<\alpha \\ 0 & \mu\geq\alpha\end{cases} \qquad \lambda>0, \quad \mu<\alpha.$$

بنا بر این، بر اساس قضیه‌ی بیز توزیع پسین متناسب است با

$$\pi\left(\alpha\middle|x\right)\propto\alpha^{n}e^{-\alpha\left(\lambda+\sum_{i=1}^{n}Lnx_{i}\right)}\underset{(\mu,\infty)}{I}(\alpha),$$

و با فرض $t^{*}=\lambda+\sum_{i=1}^{n}LnX_{i}$، می‌توان ثابت کرد توزیع پسین برابر است با

$$\pi\left(\alpha\middle|\underset{\sim}{x}\right)=\frac{\alpha^{n}e^{-(\alpha-\mu)t^{*}}\underset{(\mu,\infty)}{I}(\alpha)}{\int_{0}^{\infty}(\eta+\mu)^{n}e^{-t^{*}}\eta d\eta}.$$

لازم به ذکر است در مخرج، تغییر متغیر $\alpha=\eta+\mu$ انجام گرفته است. پس از تعیین توزیع پسین، باید براوردی برای پارامتر α بدست آوریم. اگر تابع زیان توان دوم خطاها اختیار شود، میانگین توزیع پسین به‌عنوان براوردگر بیز برای α در نظر گرفته می‌شود. بنا بر این داریم

$$\hat{\alpha}_{\beta}=E\left(\alpha\middle|\underset{\sim}{\chi}\right)=\frac{\int_{0}^{\infty}(\eta+\mu)^{n+1}e^{-t^{*}\eta}d\eta}{\int_{0}^{\infty}(\eta+\mu)^{n}e^{-t^{*}\eta}d\eta}. \qquad (۴)$$

وقتی $\mu=0$ باشد در آن صورت براوردگر بیز برای پارامتر α برابر است با:

$$\hat{\alpha}_{\beta}=\frac{n+1}{t^{*}}.$$

از آنجایی که می‌دانیم میانگین توزیع پارتو برابر است با $\frac{\alpha\delta}{\alpha-1}$، پس با این فرض که $\delta=1$، در آن صورت می‌توان اثبات نمود براورد میانگین توزیع پارتو به شیوه‌ی بیز عبارت است از

$$\hat{\mu}_{\beta} = E_{\alpha|X}\left[\frac{\alpha}{\alpha-1}\right] = \int_{\mu}^{\infty}\frac{\alpha}{\alpha-1}\Pi\left(\alpha|x\right)d\alpha$$
$$= \frac{\int_{0}^{\infty}\frac{\left(\eta+\mu\right)^{n+1}}{\eta+\mu-1}e^{-t^{*}\eta}d\eta}{\int_{0}^{\infty}\left(\eta+\mu\right)^{n}e^{-t^{*}\eta}d\eta}.$$

البته مقدار انتگرال در صورت و مخرج کسر یک فرم بسته نمی‌باشد و در نتیجه نمی‌توان آن را به شکل معمول حل نمود، لذا به‌صورت عددی می‌توان مقدار تقریبی برای $\hat{\mu}_{\beta}$ محاسبه کرد.

گاهی اوقات با تغییر توزیع پیشین می‌توان نتایج دیگری به دست آورد. از آنجایی که در توزیع پارتو فضای پارامترها و مشاهدات به یکدیگر وابسته است، می‌توان به‌عنوان یک پیشنهاد اعلام نمود توزیع پیشین دو پارامتر نیز دارای این ویژگی باشد. یکی از توزیع‌هایی که می‌توان در نظر گرفت، توزیع گامای بریده شده است، که چگالی آن به شکل زیر می‌باشد:

$$\Pi\left(\alpha\right) = \frac{\left(\alpha-\mu\right)^{\alpha-1}}{\Gamma\left(a\right)}b\,e^{-b(\alpha-\mu)}, \qquad b>0,\ \alpha>\mu,\ \alpha>0.$$

لذا بر اساس قضیه‌ی بیز توزیع پسین α به شرط مقادیر x عبارت است از

$$\Pi\left(\alpha|x\right) = \frac{e^{-\alpha\left(b+\sum_{i=1}^{n}Lnx_i\right)}\alpha^{n}\left(\alpha-\mu\right)^{a-1}}{\int_{\mu}^{\infty}e^{-\alpha\left(b+\sum_{i=1}^{n}Lnx_i\right)}\alpha^{n}\left(\alpha-\mu\right)^{a-1}d\alpha}I_{(\mu,\infty)}\left(\alpha\right). \tag{۵}$$

همان‌طور که در (۵) دیده می‌شود یک فرم پیچیده برای توزیع پسین به دست آمده است، به‌طوری که برای محاسبه‌ی میانگین، واریانس و هر پارامتر دیگر این توزیع، نمی‌توان به‌صورت معمول عبارتی را محاسبه کرد. بلکه باید با استفاده از یک نرم‌افزار آماری مانند S-Plus و یا Mathematica مقدار احتمال توزیع پسین و براورد پارامترها را با خطای بسیار ناچیز که از نظر آماری معنی‌دار نباشد تعیین کرد.

اگر بخواهیم تحت تابع زیان توان دوم خطاها، براورد پارامتر α به شیوه‌ی بیز را به دست آوریم کافی است گشتاور اول توزیع پسین را محاسبه کنیم. در این باره میانگین توزیع پسین عبارت است از

$$\hat{\alpha}_{\beta}=\frac{\int_{\circ}^{\infty} e^{-\eta\left(b_{\circ}+\sum_{i=1}^{n} Lnx_i\right)}(\eta+\mu)^{n+1}\,\eta^{a_{\circ}-1}d\eta}{\int_{\circ}^{\infty} e^{-\eta\left(b_{\circ}+\sum_{i=1}^{n} Lnx_i\right)}(\eta+\mu)^{n}\,\eta^{a_{\circ}-1}d\eta}. \qquad (۶)$$

همان‌طور که دیده می‌شود در این حالت نیز وقتی پارامتر δ معلوم باشد محاسبات مربوط به تعیین براورد بیز α به انتگرال‌هایی منجر می‌گردد که فقط به‌صورت عددی قابل محاسبه است. این موضوع وقتی هر دو پارامتر α و δ مجهول باشند، بسیار مشکل‌تر خواهد بود. لذا می‌توان گفت، در بسیاری موارد (از جمله مورد قبل) هرچند به روش بیز کارایی براوردکننده‌ها بیش‌تر است، اما گاهی اوقات پیچیدگی محاسبات باعث می‌گردد از ورود به این حوزه خودداری شود.

۳- تحلیل بیزی پارامترهای توزیع پارتو وقتی هر دو پارامتر توزیع مجهول باشد

وقتی هر دو پارامتر مجهول باشند، توزیع پسین مربوط به چگالی پارتو را نیز می‌توان بررسی نمود. اما طبیعی است این مطلب بر توزیع‌های پیشین نیز تأثیرگذار خواهد بود. مالیک (۱۹۷۰) یک خانواده از توزیع‌های پیشین برای توزیع پارتو در نظر گرفت که در آن فرض وابستگی پارامترهای α و δ به یکدیگر نیز وجود داشت [۷]. در اینجا نیز این وابستگی مورد بررسی قرار می‌گیرد، زیرا در کاربرد، معمولاً بین این دو پارامتر وابستگی معنی‌داری وجود دارد.

فرض کنید X_1 ، ... ، X_n ، نمونه‌ی تصادفی n تایی از توزیع پارتو با پارامتر مقیاس δ و پارامتر شکل α باشند. همچنین فرض کنید توزیع‌های پیشین α و δ به ترتیب توزیع‌های نمایی و پارتو به شکل زیر باشد:

$$\alpha \sim E\left(\lambda, \mu\right),$$
$$\delta \sim P\left(\beta, \delta_{\circ}\right).$$

با توجه به این که تابع درست‌نمایی و توزیع‌های پیشین در دسترس می‌باشد، می‌توان توزیع پسین را محاسبه نمود. از آنجایی که می‌دانیم توزیع پسین متناسب است با حاصل‌ضرب تابع درست‌نمایی و توزیع پیشین، بنا بر این می‌توان نوشت

$$\Pi\left(\alpha, \delta \middle| x\right) \propto \alpha^{n} \delta^{n\alpha+\beta-1} e^{-\alpha\left(\sum_{i=1}^{n} \mathrm{Ln} X_{i} + \lambda\right)}. \qquad (۷)$$

رابطه‌ی (۷) را می‌توان با یک ضریب تناسب به یک تساوی تبدیل نمود. زیرا توزیع پسین به‌عنوان یک چگالی و انتگرال آن مساوی یک می‌باشد. به‌سادگی می‌توان گفت مقدار ثابت که به تابع حاشیه‌ای x معروف است برابر است با

$$m\left(x\right) = \int_{\mu}^{\infty} \int_{\circ}^{\infty} \alpha^{n} \delta^{n\alpha+\beta-1} e^{-\alpha\left(\sum_{i=1}^{n} \mathrm{Ln} X_{i} + \lambda\right)} d\delta\, d\alpha.$$

با تعریف $t^{*} = \lambda + \sum_{i=1}^{n} LnX_{i}$ و $V_{\circ} = \min(\delta_{1}, X_{1})$ خواهیم داشت:

$$m\left(x\right) = V_{\circ}^{\beta} \int_{\mu}^{\infty} \frac{\alpha^{n}}{n\alpha+\beta} e^{-\alpha\left(t^{*} - nLnV_{\circ}\right)} d\alpha.$$

با تغییر متغیر $\eta = \alpha - \mu$، تعریف $\beta^{*} = \frac{\rho}{n}$ و $t^{**} = t^{*} - nLnV_{\circ}$، توزیع حاشیه‌ای به فرم ساده‌تری تبدیل می‌شود و در نتیجه بر اساس آن می‌توان توزیع پسین را به شکل ساده‌تری تبدیل نمود.

پس از تعیین توزیع پسین اگر بخواهیم به‌طور جداگانه توزیع‌های حاشیه‌ای α و δ را محاسبه نماییم، می‌توان با انتگرال‌گیری از توزیع پسین نسبت به تولید توزیع‌های حاشیه‌ای α و یا δ اقدام نمود. پس از تهیه‌ی توزیع‌های پسین تک‌متغیره مربوط به α و δ، می‌توانیم براورد به روش بیز این پارامترها را به دست آوریم. طبیعی است در این بخش نیز باید با استفاده از روش‌های عددی این عمل را انجام داد. نتایج مشابه آنچه در بخش قبل مطرح شد، می‌باشد و از ذکر دوباره‌ی آن‌ها پرهیز می‌گردد.

برای به دست آوردن براورد پارامترها به روش بیز اولین سئوال این است که بر اساس چه تابع زیانی می‌خواهیم این براورد را به دست آوریم. معمولاً مهم‌ترین تابع زیان، تابع زیان توان دوم خطاها می‌باشد. اگر از این تابع زیان که اختلاف بین پارامتر و براوردگر را به‌صورت توان دوم نمایش می‌دهد استفاده نماییم در آن صورت، میانگین توزیع پسین به‌عنوان یک براوردگر مورد استفاده قرار می‌گیرد، که این میانگین برای α و δ به ترتیب برابر است با

$$\hat{\alpha}_{\beta} = E\left[\alpha \middle| x\right] = \frac{\int_{\circ}^{\infty} \frac{(\eta+\mu)^{n+1}}{\eta+\mu+\beta^{*}} e^{-\eta t^{**}} d\eta}{\int_{\circ}^{\infty} \frac{(\eta+\mu)^{n}}{\eta+\mu+\beta^{*}} e^{-\eta t^{**}} d\eta} \tag{۸}$$

و

$$\hat{\delta}_{\beta} = E\left[\delta \middle| x\right] = \frac{v_{\circ}\int_{\circ}^{\infty} \frac{(\eta+\mu)^{n}}{\eta+\mu+\frac{\beta+1}{n}} e^{-\eta t^{**}} d\eta}{\int_{\circ}^{\infty} \frac{(\eta+\mu)^{n}}{\eta+\mu+\beta^{*}} e^{-\eta t^{**}} d\eta}.$$

که در آن

$$\beta^{*} = \frac{\beta}{n}, \qquad t^{**} = \sum_{i=1}^{n} \ln X_{i} + \lambda - n \ln v_{\circ}.$$

نکته‌ی قابل ذکر این است که در اینجا فرض شده است پارامترهای مربوط به توزیع‌های پیشین مقادیر معلومی دارند. اما در عمل معمولاً این مقادیر نیز مجهول می‌باشند، لذا استفاده از این روش در عمل باعث می‌شود تا ابتدا برای پارامترهای اولیه نیز براوردی بدست آورده سپس بر مبنای آن، پارامترهای مورد مطالعه را براورد کنیم. این نظریه که به نظریه‌ی بیز تجربی[۶] معروف است دارای یک روش علمی خاص است که در کاربرد بسیار مهم می‌باشد و می‌تواند تحلیل‌های بیزی را تحت تأثیر قرار دهد.

برای گریز از این موضوع یک راه اصولی، تعیین براوردی برای پارامترهای اولیه با استفاده از مشاهدات می‌باشد. البته بسیاری از آمارشناسان استفاده از مشاهدات را در تعیین براورد پارامترهای اولیه مناسب نمی‌دانند، ولی از آنجایی که می‌خواهیم توافقی بین توزیع‌های پیشین و تابع درست‌نمایی مشاهدات انجام دهیم، این توافق با استفاده از ادغام این دو نظریه و آن هم به‌وسیله‌ی براورد پارامترها می‌تواند انجام شود. گاهی اوقات می‌توان بر اساس پیش‌فرض اولیه، پارامترهای توزیع‌های پیشین را معلوم فرض کرد، سپس تحلیل بیزی را برای پارامترهای توزیع پارتو بکار برد. هر دو حالت در کاربرد از جایگاه معتبری برخوردار است. در قسمت بعد با ذکر یک مثال کاربردی و ارائه‌ی راهکارهای مناسب براورد مربوط به پارامترها و در نتیجه توزیع پسین براوردگرها را می‌توان ارائه نمود.

۴- کاربرد

امروزه یکی از موضوعاتی که باید در جامعه مورد توجه قرار گیرد مسئله‌ی درامد خانوارها و عوامل مؤثر بر آن می‌باشد. آنچه مسلم است در یک پژوهش نمی‌توان همه عواملی را که ممکن است بر متغیر پاسخ تأثیرگذار باشند شناسایی کرد و مورد ارزیابی قرار داد، اما به‌وسیله‌ی مدل‌های آماری تلاش می‌شود، با کم‌ترین خطا بهترین مدل را برای تجزیه و تحلیل داده‌ها مورد استفاده قرار داد. اکنون برای بکارگیری مباحث نظری بخش‌های قبل، اطلاعات جمع‌آوری‌شده در سال ۱۳۸۳ در مورد وضعیت خانوارهای تهران، درامد خانوارها و عوامل مؤثر بر آن نظیر سن، جمعیت، وضعیت تأهل و ... برای فرد شاغل خانوارها در نظر گرفته شده است. در این مطالعه جامعه‌ی آماری شامل نام خانوارهای معمولی ساکن در نقاط شهری تهران است. خانوارهای معمولی غیر ساکن و دسته‌جمعی از جامعه‌ی این طرح مستثنی است. چارچوب آماری طرح، شامل فهرست تمام بلوک‌های موجود در نقاط شهری به انضمام تعداد خانوارها در هر یک از بلوک‌ها می‌باشد.

۴-۱- طرح نمونه‌گیری

از آنجا که دسترسی به همه‌ی بلوک‌های موجود پرهزینه و زمان‌بر است، به‌وسیله‌ی یک طرح نمونه‌گیری خوشه‌ای دومرحله‌ای نمونه‌ای به اندازه‌ی ۲۵۵ خانوار انتخاب نموده‌ایم.

ابتدا تعداد خانوار نمونه در هر استان را به ۵ عدد (تعداد خانوار نمونه مورد نظر در هر بلوک) تقسیم کرده تا تعداد بلوک نمونه بدست آید. سپس به روش نمونه‌گیری با احتمال متناسب با اندازه و به‌صورت سیستماتیک خطی با گام ۳، بلوک‌های نمونه را انتخاب نموده‌ایم. اگر بلوک نمونه کم‌تر از ۵ خانوار داشته باشد بلوک یا بلوک‌های دیگری در داخل همان حوزه انتخاب و به بلوک نمونه ضمیمه می‌شود، تا جمع تعداد خانوار به حد اقل مورد نیاز برسد. پس از تعیین حجم نمونه‌ی لازم پرسش‌های پیرامون موارد زیر مطرح شد:

سواد: هر عضو ۶ ساله و بیش‌تر که متن فارسی را به فارسی یا هر زبان دیگر بتواند بخواند و بنویسد، باسواد تلقی می‌شود که در دو سطح ۱- باسواد، ۲- بی‌سواد، تقسیم‌بندی گردید.

وضع فعالیت: وضع فعالیت در یکی از حالت‌های زیر در نظر گرفته شد:
۱- شاغل، ۲- بیکار، ۳- دارای درامد بدون کار، ۴- محصل، ۵- خانه‌دار و ۶- سایر.

جنسیت: ۱- مرد، ۲- زن.

وضع تأهل: ۱- دارای همسر، ۲- بی‌همسر بر اثر فوت همسر، ۳- بی‌همسر بر اثر طلاق و ۴- مجرد.

درامد: درامد شامل تمام وجوه و ارزش کالایی است که در برابر کار انجام‌شده یا سرمایه بدست آمده است و یا با استفاده از منابع دیگر در زمان کاری مورد نظر به خانواده تعلق گرفته است.

خط فقر: حد اقل درامد سالیانه‌ی مورد نیاز برای زندگی می‌باشد.

از آنجا که هدف از این پژوهش ارائه‌ی مبانی نظری و مطالعه‌ی براوردهای پارامترهای توزیع‌های پیشین می‌باشد لذا خانوارهای استان مرکزی مورد بررسی قرار گرفتند اما به‌طور مشابه می‌توان برای سایر استان‌ها نیز این تحقیق را انجام داد. در آغاز ذکر چند موضوع دارای اهمیت است.

الف- درامدها سالیانه می‌باشند.

ب- برای خانوارهای داخل نمونه که در آن‌ها چند فرد صاحب درامد بوده‌اند، مجموع درامدهایی که افراد آن خانوار کسب کرده‌اند به‌عنوان درامد خانوار محسوب می‌شود.

ج- برای سهولت محاسبات، واحد مربوط به مقدار درامد، میلیارد ریال در نظر گرفته شده است.

د- بر اساس اطلاعات کسب‌شده از سازمان مدیریت و برنامه‌ریزی، خط فقر در مناطق مختلف شهری کشور برای یک خانوار ۵ نفره‌ی شهری بر اساس درامد سالیانه رقم ۱۲۴۸۶۸۰ ریال می‌باشد.

با توجه به اطلاعات جمع‌آوری‌شده از خانوارهای نمونه‌ای نام‌برده جدول ۱ درامد این افراد را نشان می‌دهد. همان‌طور که در این جدول مشخص است، برحسب درامدهای مختلف، کل خانوارها را نمی‌توان در یک طبقه قرار داد و آن‌ها را به ۹ طبقه تقسیم نموده‌ایم. البته فراوانی مشاهدات کاملاً با یکدیگر متفاوت است. که این موضوع باعث می‌شود تا تغییرات در تمام رده با یکدیگر به‌طور یکسان نباشد. در مورد جدول ۱، ابتدا باید بررسی نمود که آیا می‌توان ادعا کرد این داده‌ها از توزیع پارتو تبعیت می‌کند؟ در واقع باید آزمون زیر را بیازماییم.

$$H_{\circ}: Y \sim pa(\alpha,\delta)$$
$$H_{\backprime}: \text{درست نیست.}\ H_{\circ}$$

میانگین و میانه‌ی داده‌ها در جدول ۱ به ترتیب برابر است با ۰/۰۲۲۷ و ۰/۰۱۹۵۹. مقدار آماره‌ی $\chi^{۲}$ نیز برابر با ۹۴/۸۷۸۵ است که مقدار متناظر آن در جدول $\chi^{۲}_{(۶)} = ۱۴/۴$ است. پس فرض $H_{\circ}$ مورد پذیرش قرار نمی‌گیرد. برای این منظور $\exp(x)$ داده‌ها و سپس مقدار آماره $\chi^{۲}$ محاسبه گردید که در این حالت $\chi^{۲} = ۷/۷۳$ بدست آمد و این موضوع حاکی از آن است که داده‌های به‌صورت $\exp(X)$ از توزیع پارتو تبعیت می‌کنند. اکنون با استفاده از تابع ماکسیمم درست‌نمایی مربوط به توزیع پارتو، براورد δ,α را می‌توان محاسبه نمود که عبارت‌اند از: $\hat{\alpha} = ۶۶/۹۷۳۶$ و $\hat{\delta} = ۱/۰۰۷۸$.

محاسبات همچنین نشان می‌دهد که ضریب جینی$(\hat{g})$ برابر با ۰/۰۰۷۵۲، شاخص فقر$(\hat{\alpha}_{\circ})$ برابر با ۳۸% و احتمال بقاء برابر با ۶۲% است. از این نتایج می‌توان نتیجه گرفت:

الف- توزیع درامد در نمونه‌ی مورد مطالعه عادلانه است.

ب- کم‌تر از ۵۰ درصد افراد نمونه دارای درامدی کم‌تر از حد اقل درامد مورد نیاز برای زندگی هستند.

جدول ۱- فراوانی مشاهدات برحسب درامدهای مختلف

طبقه	حد پایین طبقه	حد بالای طبقه	فراوانی (a_i)
۱	۰/۰۰۷۷۹۹۹۹۹۵	۰/۰۱۸۷۹۹۹۹۹۵	۱۲۰
۲	۰/۰۱۸۷۹۹۹۹۹۵	۰/۰۲۹۷۹۹۹۹۹۵	۸۴
۳	۰/۰۲۹۷۹۹۹۹۹۵	۰/۰۴۰۷۹۹۹۹۹۵	۳۳
۴	۰/۰۴۰۷۹۹۹۹۹۵	۰/۰۵۱۷۹۹۹۹۹۵	۸
۵	۰/۰۵۱۷۹۹۹۹۹۵	۰/۰۶۲۷۹۹۹۹۹۵	۴
۶	۰/۰۶۲۷۹۹۹۹۹۵	۰/۰۷۳۷۹۹۹۹۹۵	۳
۷	۰/۰۷۳۷۹۹۹۹۹۵	۰/۰۸۴۷۹۹۹۹۹۵	۱
۸	۰/۰۸۴۷۹۹۹۹۹۵	۰/۰۹۵۷۹۹۹۹۹۵	۱
۹	۰/۰۹۵۷۹۹۹۹۹۵	-	۱

اما همان‌طور که مبانی نظری ارائه‌شده بیان می‌کند قصد داریم تا تحت پیشینهای در نظر گرفته‌شده در این پژوهش و بر اساس تابع زیان توان دوم خطاها، براورد توزیع‌های پسین و براورد پارامتر مورد نیاز را به دست آوریم. جدول ۲ وضعیت این براوردگرها را تحت پیشین‌های مختلف معرفی می‌کند.

با توجه به براوردهای بدست آمده برای پارامترهای اولیه، توزیع‌های پسین به شکل زیر معرفی می‌شود:

تحت نمایی:

$$\pi\left(\alpha,\delta\middle|x\right)=\frac{\alpha^{۲۵۵}}{۰/۹۸۶۴۳۲}\delta^{۲۵۵\alpha-۰/۵}e^{-\alpha(۶/۵۹۶)}\underset{(۰,۰/۷۱۸)}{I}\left(\delta\right)\underset{(۰/۷۴,\infty)}{I}\left(\alpha\right)$$

تحت گاما:

$$\pi\left(\alpha,\delta\middle|x\right)=\frac{\alpha^{۲۵۶}}{۰/۹۹۵۲۲۵}\delta^{۲۵۵\alpha-۰/۹}e^{۶/۲۹۶\alpha}\underset{(۰,۰/۷۱۶)}{I}\left(\delta\right)$$

جدول ۲- تعیین براورد پارامترهای به روش نظریه‌ی بیز

براورد پارامتر تحت تابع زیان توان دوم خطاها	براورد پارامترهای اولیه	توزیع پیشین α
$\hat{\alpha}_{\beta} = ۲/۷۹۹۹۲$	$\hat{\lambda} = ۰/۸$	
$\hat{\delta}_{\beta} = ۰/۷۱۶۹۹۳$	$\hat{\mu} = ۰/۷۴$	
$\hat{g}_{\beta} = ۰/۲۱۸۶۷۲$	$\hat{\delta}_{\circ} = ۰/۷۱۸$	نمایی دوپارامتری
$\hat{\mu}_{\beta} = ۱/۱۱۹۱۸$	$\hat{\beta} = ۰/۵$	
$\hat{\alpha}_{\circ\beta} = ۰/۶۲۱۵$		
$\hat{\lambda}_{\circ\beta} = ۰/۳۷۸۴$		
$\hat{\alpha}_{\beta} = ۲/۷۹۸۲۶$	$\hat{\alpha}_{\circ} = ۲$	
$\hat{\delta}_{\beta} = ۰/۷۱۴۹۹۴$	$\hat{b}_{\circ} = ۰/۵$	
$\hat{g}_{\beta} = ۰/۱۵۶۶۷۹$	$\hat{\delta}_{\circ} = ۰/۷۱۶$	گاما
$\hat{\mu}_{\beta} = ۱/۱۱۶۴۲$	$\hat{\beta} = ۰/۱$	
$\hat{\alpha}_{\circ\beta} = ۰/۶۲۴۲۴$		
$\hat{\lambda}_{\circ\beta} = ۰/۳۷۵۷۵$		

ملاک مقایسه:

یک ملاک مقایسه برای عملکرد شیوه‌ی فراوانی‌گرا و بیز می‌تواند واریانس‌های براوردکننده‌ها باشد. با توجه به نتایج جدول ۳ می‌توان گفت چون مقدار براورد واریانس به دست آمده برای براورد پارامترها با استفاده از روش بیز کوچک‌تر از براورد واریانس به دست‌آمده در حالت فراوانی‌گرا است لذا روش بیز بهتر عمل نموده است. همچنین در میان روش‌های بیز با توجه به این که براورد واریانس δ و α تحت توزیع پیشین گاما کم‌تر است پس می‌توان نتیجه گرفت در این حالت نتایج بهتری بدست آمده است.

جدول ۳- براورد واریانس براوردگرها

براورد واریانس δ	براورد واریانس α	
۰/۵۹ $\times$۱۰$^{-۵}$	۰/۸۰۸۲	روش فراوانی‌گرا
۱/۰۹۹۸$\times$۱۰$^{-۶}$	۰/۰۳۰۷۴۴	روش بیز با توزیع پسین نمایی
۱/۰۱۶۷ $\times$۱۰$^{-۶}$	۰/۰۳۰۵۸	روش بیز با توزیع پیشین گاما

جدول ۴- جدول تحلیل واریانس مربوط به تأثیر عوامل کمکی بر درامد خانوار

معنی‌داری	مقدار F	مجموع مربعات	درجه‌ی آزادی	عوامل
۰/۳	۳	۸۵۶	۱	جنسیت
۰/۰۰۱	۱۶/۵	۸۴۵۳	۱	سواد
۰/۱۹	۵/۴	۷۶۳۳	۵	وضعیت فعالیت
۰/۲۴	۲/۵	۵۶۳۵	۳	وضعیت تأهل
۰/۰۰۲	۱۱/۵	۴۳۵۲	۴	تعداد افراد شاغل
			۲۴۰	خطا

بررسی عوامل مؤثر بر درامد خانوارهای نمونه

نتایج حاصل در خصوص بررسی عوامل مؤثر جنسیت، وضع سواد، وضعیت فعالیت، وضعیت تأهل و تعداد افراد شاغل بر عامل درامد خانوار بر اساس مدل های ارائه شده در بخش‌های قبل در جدول ۴ آورده شده است که نشان می‌دهد دو عامل تعداد افراد شاغل و وضعیت سواد بر درامد خانوارها بسیار مؤثرند. بررسی بیش‌تر بیان می‌کند، هر چه وضعیت سواد بهتر و تعداد افراد شاغل افزایش یابد به‌طور چشم‌گیری درامد خانوار نیز افزایش می‌یابد. پس مدل پیشنهادی مدلی است که در آن بر اساس عوامل پیشنهادی، اثر دو عامل تعداد افراد شاغل و وضعیت سواد بر درامد خانوارها مدل‌بندی گردد.

نتیجه گیری

می‌توان گفت در این مطالعه روش بیزی پیشنهاد شده بسیار کارا تر از روش فراوانی‌گرا عمل نموده است. البته می توان این مطالعه را در حالتی که مدل غیرخطی است، انجام داد که این کار را به یک پژوهش دیگر می‌توان ارجاع نمود.

توضیحات

۱. Frequentist approach
۲. Extereme Values
۳. Sufficient Statistics
۴. Frequentist Approach
۵. Conjugate Prior
۶. Empirical Bayes

مرجع‌ها

[1] Arnold, B.C. and Press, S.J. (1983). Estimation of parameters for Pareto data. *Journal of Economics*, **21**, 286-306.

[2] Champernowne, D.J. (1953). A model of income distributions. *Economics Journal*, **63**, 318-351.

[3] Galambos, J. (1994). *The Asymptotic Theory of Extreme Order Statistics* Wiley, New York.

[4] Genedenko, B. (1943). Surla distribution limite du treme maximum dune serie a leatorie. *Annals of Mathematics*, **44**, 423-453.

[5] Jakinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, **81**, 158-171.

[6] Levy, M., Solomon, S. (1997). New Evidence for the Pareto-Law Distribution of Wealth, *Physica A*, **242**, 90-94.

[7] Malik, H.J. (1970). Estimation of the parameters of the Pareto Distribution, *Metrika*, **15**, 126-132.

[8] Mardina, K.V. (1962). Multivariate Pareto distributions. *Ann. of Math. Stat*, **33**, 1008-1015.

[9] Pareto, V. (1964). Cours d'Economie Politique: Nouvelle edition par G.-H. Bousquet et G. Busino, Librairie Droz, Geneva. 299-345.

[10] Persky, J. (1992). Retrospectives: Pareto's law. *Journal of Economics perspective*, **6**, 181–192.

[11] Wingo, D.R. (1982). Unimodality of the Pareto distribution likelihood function for multi censored samples and implications for estimation. *Common Statistics. Theory and Methods*, **11**, 1129–1138.

فرزاد اسکندری
دکتری آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه‌طباطبایی، گروه آمار.
رایانشانی: ffeskandari@yahoo.com

رقیه قبادی
فوق لیسانس آمار
تهران، وزارت بازرگانی، اداره‌ی آمار و برنامه‌ریزی.