

تحلیل بیزی توزیع درآمد خانوار

فرزاد اسکندری^{†*} و رقیه قبادی[‡]

[†] دانشگاه علامه طباطبائی

[‡] وزارت بازرگانی

چکیده: مطالعه‌ی وضعیت برآورد 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 برآورد 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 برآورد پارامترها با استفاده از توزیع پسین به دست آمده است، با محاسبه‌ی واریانس‌های برآوردگرها، 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؛ مدل بندی 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، به طوری که می‌توان به کارهایی که گالامبوس [۳]، گندنکو [۴] و جکینسون [۵] انجام داده‌اند اشاره نمود. اما اولین فردی که به طور رسمی این توزیع را معرفی کرد ویلفردو پارتو (۱۹۶۴)، اقتصاددان مشهور ایتالیایی بوده است که این توزیع را به عنوان توزیع درآمد بیان نمود [۹]. همچنین چمپرنون ثابت کرد که توزیع درآمد، صرف نظر از توزیع اصلی آن می‌تواند با توزیع پارتو تقریب شود [۲]. در خصوص مطالعه‌ی توزیع درآمد و ارتباط آن با توزیع آماری پارتو، افراد زیادی مطالعه نمودند که از آن جمله می‌توان به پرسکه [۱۰]، له‌وی و سولومون [۶] اشاره نمود. قابل توجه 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 برآورد پارامترها بدست آورده است [۸]. همچنین برای بررسی برآوردگرهای ماکسیمم درست‌نمایی پارامترهای بهین توزیع، می‌توان به مقاله وینگو اشاره کرد [۱۱]. تحلیل بیزی برآورد 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 به عنوان 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^۵ برای پارامترهای مدل، به تحلیل بیزی توزیع پارتو پرداخته، آن را در یک مثال واقعی آن را به کار برده و نتایج حاصل را به وسیله‌ی نسبت واریانس‌ها مقایسه می‌کنیم. این مقاله شامل ۴ بخش می‌باشد. در بخش دوم این مقاله به تحلیل بیزی توزیع درآمد 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 δ معلوم می‌باشد.

فرض کنید X_1, \dots, X_n یک نمونه‌ی تصادفی از تابع توزیع پارتو به صورت زیر باشند:

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

به‌سادگی می‌توان مشخص نمود که چگالی آن برابر است با

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

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

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \ln X_i - n \ln \hat{\delta}}, \quad \hat{\delta} = X_{(1)}.$$

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

پس از تعیین برآوردهای ماکسیمم درست‌نمایی پارامترها می‌توانیم با استفاده از بازه‌های اطمینان مناسب یک برآورد 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 برآوردهای کمیت محوری حاصل می‌شود، χ^2 (۲) درجه آزادی می‌باشد که در (۲) آمده است.

$$(2) \quad 2\alpha \left[\sum_{i=1}^n \ln X_i - n \ln X_{(1)} \right] \sim \chi^2(2n-2)$$

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

$$P \left[X_{\frac{r}{\alpha}}^{\alpha} (\alpha n - \alpha) < \alpha \left(\sum_{i=1}^n \text{Ln} X_i - n \text{Ln} X_{(1)} \right) < X_{\frac{1-r}{1-\alpha}}^{\alpha} (\alpha n - \alpha) \right] = 1 - r \quad 0 \leq r \leq 1$$

و در نتیجه بازه‌ی اطمینان برای پارامتر α عبارت است از

$$(3) \quad \left(\frac{\chi_{\frac{r}{\alpha}}^{\alpha} (\alpha n - \alpha)}{\alpha \left(\sum_{i=1}^n \text{Ln} X_i - n \text{Ln} X_{(1)} \right)}, \frac{\chi_{\frac{1-r}{1-\alpha}}^{\alpha} (\alpha n - \alpha)}{\alpha \left(\sum_{i=1}^n \text{Ln} X_i - n \text{Ln} X_{(1)} \right)} \right)$$

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 برآورد کنیم، حالت‌های مختلفی بوجود می‌آید به طوری 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 $\delta = 1$ تابع درست‌نمایی عبارت است از

$$L(\alpha) = \alpha^n \exp \left\{ -(\alpha + 1) \sum_{i=1}^n \text{Ln} X_i \right\}$$

اگر توزیع پیشین نمایی دوپارامتری اختیار شود، خواهیم داشت

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

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

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

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

$$\pi(\alpha|x) = \frac{\alpha^n e^{-(\alpha-\mu)t^*} I(\alpha)_{(\mu, \infty)}}{\int_0^\infty (\eta + \mu)^n e^{-t^* \eta} \eta d\eta}.$$

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

$$(۴) \quad \hat{\alpha}_\beta = E(\alpha|x) = \frac{\int_0^\infty (\eta + \mu)^{n+1} e^{-t^* \eta} d\eta}{\int_0^\infty (\eta + \mu)^n e^{-t^* \eta} d\eta}.$$

وقتی $\mu = 0$ باشد در آن صورت برآوردگر 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 برآورد میانگین توزیع پارتو به شیوه‌ی بیز عبارت است از

$$\hat{\mu}_\beta = E_{\alpha|x} \left[\frac{\alpha}{\alpha-1} \right] = \int_{\mu}^{\infty} \frac{\alpha}{\alpha-1} \Pi(\alpha|x) d\alpha$$

$$= \frac{\int_0^{\infty} \frac{(\eta+\mu)^{n+1}}{\eta+\mu-1} e^{-t^* \eta} d\eta}{\int_0^{\infty} (\eta+\mu)^n e^{-t^* \eta} d\eta}.$$

البته مقدار انتگرال در صورت و مخرج کسر یک فرم بسته نمی‌باشد و در نتیجه نمی‌توان آن را به شکل معمول حل نمود، لذا به صورت 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 به عنوان 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:

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

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

$$(5) \quad \Pi(\alpha|x) = \frac{e^{-\alpha \left(b + \sum_{i=1}^n Lnx_i \right)} \alpha^n (\alpha-\mu)^{\alpha-1}}{\int_{\mu}^{\infty} e^{-\alpha \left(b + \sum_{i=1}^n Lnx_i \right)} \alpha^n (\alpha-\mu)^{\alpha-1} d\alpha} I(\alpha).$$

همان‌طور که در (5) دیده می‌شود یک فرم پیچیده برای توزیع پسین به دست آمده است، به طوری که برای محاسبه‌ی میانگین، واریانس و هر پارامتر دیگر این توزیع، نمی‌توان به صورت معمول عبارتی را محاسبه کرد. بلکه باید با استفاده از یک نرم‌افزار آماری مانند S-Plus و یا Mathematica مقدار احتمال توزیع پسین و برآورد پارامترها را با خطای بسیار ناچیز که از نظر آماری معنی‌دار نباشد تعیین کرد.

اگر بخواهیم تحت تابع زیان توان دوم خطاها، برآورد 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

$$(۶) \quad \hat{\alpha}_\beta = \frac{\int_0^\infty e^{-\eta \left(b_0 + \sum_{i=1}^n Lnx_i \right)} (\eta + \mu)^{n+1} \eta^{a_0-1} d\eta}{\int_0^\infty e^{-\eta \left(b_0 + \sum_{i=1}^n Lnx_i \right)} (\eta + \mu)^n \eta^{a_0-1} d\eta}.$$

همان‌طور که دیده می‌شود در این حالت نیز وقتی پارامتر δ معلوم باشد محاسبات مربوط به تعیین برآورد بیز α به انتگرال‌هایی منجر می‌گردد که فقط به صورت 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 برآوردکننده‌ها 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, \dots, 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:

$$\alpha \sim E(\lambda, \mu),$$

$$\delta \sim P(\beta, \delta_0).$$

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 حاصل ضرب تابع درست‌نمایی و توزیع پیشین، بنا بر این می‌توان نوشت

$$(V) \quad \Pi(\alpha, \delta | x) \propto \alpha^n \delta^{n\alpha + \beta - 1} e^{-\alpha \left(\sum_{i=1}^n \text{Ln} X_i + \lambda \right)}.$$

رابطه‌ی (V) 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

$$m(x) = \int_{\mu}^{\infty} \int_{\delta_0}^{\infty} \alpha^n \delta^{n\alpha + \beta - 1} e^{-\alpha \left(\sum_{i=1}^n \text{Ln} X_i + \lambda \right)} d\delta d\alpha.$$

با تعریف $t^* = \lambda + \sum_{i=1}^n \text{Ln} X_i$ و $V_0 = \min(\delta, X_i)$ خواهیم داشت:

$$m(x) = V_0^\beta \int_{\mu}^{\infty} \frac{\alpha^n}{n\alpha + \beta} e^{-\alpha(t^* - n \text{Ln} V_0)} d\alpha.$$

با تغییر متغیر $\eta = \alpha - \mu$ ، تعریف $\beta^* = \frac{\rho}{n}$ و $t^{**} = t^* - n \text{Ln} V_0$ ، 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 برآورد 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 برآورد پارامترها به روش بیز اولین سؤال این است که بر اساس چه تابع زیانی می‌خواهیم این برآورد را به دست آوریم. معمولاً مهم‌ترین تابع زیان، تابع زیان توان دوم خطاها می‌باشد. اگر از این تابع زیان که اختلاف بین پارامتر و برآوردگر را به صورت توان دوم نمایش می‌دهد استفاده نماییم در آن صورت، میانگین توزیع پسین به‌عنوان یک برآوردگر مورد استفاده قرار می‌گیرد، که این میانگین برای α و δ به ترتیب برابر است با

$$(۸) \quad \hat{\alpha}_\beta = E[\alpha|x] = \frac{\int_0^\infty \frac{(\eta + \mu)^{n+1}}{\eta + \mu + \beta^*} e^{-\eta t^{**}} d\eta}{\int_0^\infty \frac{(\eta + \mu)^n}{\eta + \mu + \beta^*} e^{-\eta t^{**}} d\eta}$$

و

$$\hat{\delta}_\beta = E[\delta|x] = \frac{v_0 \int_0^\infty \frac{(\eta + \mu)^n}{\eta + \mu + \frac{\beta + 1}{n}} e^{-\eta t^{**}} d\eta}{\int_0^\infty \frac{(\eta + \mu)^n}{\eta + \mu + \beta^*} e^{-\eta t^{**}} d\eta}.$$

که در آن

$$\beta^* = \frac{\beta}{n}, \quad t^{**} = \sum_{i=1}^n \ln X_i + \lambda - n \ln v_0.$$

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 برآوردی بدست آورده سپس بر مبنای آن، پارامترهای مورد مطالعه را برآورد کنیم. این نظریه که به نظریه‌ی بیز تجربی 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 برآوردی برای پارامترهای اولیه با استفاده از مشاهدات می‌باشد. البته بسیاری از آمارشناسان استفاده از مشاهدات را در تعیین برآورد 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 به وسیله‌ی برآورد 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 برآورد مربوط به پارامترها و در نتیجه توزیع پسین برآوردگرها را می‌توان ارائه نمود.

۴- کاربرد

امروزه یکی از موضوعاتی که باید در جامعه مورد توجه قرار گیرد مسئله‌ی درآمد 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 به وسیله‌ی 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 جمع‌آوری شده در سال ۱۳۸۳ در مورد وضعیت خانوارهای تهران، درآمد خانوارها و عوامل مؤثر بر آن نظیر سن، جمعیت، وضعیت تأهل و ... برای فرد شاغل خانوارها در نظر گرفته شده است. در این مطالعه جامعه 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، به وسیله‌ی یک طرح نمونه‌گیری خوشه‌ای دو مرحله‌ای 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 به صورت 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 درآمد 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.

درآمد: درآمد 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 درآمد سالیانه‌ی مورد نیاز برای زندگی می‌باشد.

از آنجا که هدف از این پژوهش ارائه‌ی مبانی نظری و مطالعه‌ی برآوردهای 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- درآمدها سالیانه می‌باشند.

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

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

د- بر اساس اطلاعات کسب شده از سازمان مدیریت و برنامه ریزی، خط فقر در مناطق مختلف شهری کشور برای یک خانوار ۵ نفره شهری بر اساس درآمد سالانه رقم ۱۲۴۸۶۸۰ ریال می باشد.

با توجه به اطلاعات جمع آوری شده از خانوارهای نمونه ای نام برده جدول ۱ درآمد این افراد را نشان می دهد. همان طور که در این جدول مشخص است، برحسب درآمدهای مختلف، کل خانوارها را نمی توان در یک طبقه قرار داد و آن ها را به ۹ طبقه تقسیم نموده ایم. البته فراوانی مشاهدات کاملاً با یکدیگر متفاوت است. که این موضوع باعث می شود تا تغییرات در تمام رده با یکدیگر به طور یکسان نباشد. در مورد جدول ۱، ابتدا باید بررسی نمود که آیا می توان ادعا کرد این داده ها از توزیع پارتو تبعیت می کند؟ در واقع باید آزمون زیر را بیازماییم.

$$H_0: Y \sim pa(\alpha, \delta)$$

$$H_1: \text{درست نیست.}$$

میانگین و میانه ی داده ها در جدول ۱ به ترتیب برابر است با ۰/۰۲۲۷ و ۰/۰۱۹۵۹. مقدار آماره ی χ^2 نیز برابر با ۹۴/۸۷۸۵ است که مقدار متناظر آن در جدول $\chi^2_{(9)} = 14/4$ است. پس فرض H_0 مورد پذیرش قرار نمی گیرد. برای این منظور $\exp(x)$ داده ها و سپس مقدار آماره χ^2 محاسبه گردید که در این حالت $\chi^2 = 7/73$ بدست آمد و این موضوع حاکی از آن است که داده های به صورت $\exp(X)$ از توزیع پارتو تبعیت می کنند. اکنون با استفاده از تابع ماکسیمم درست نمایی مربوط به توزیع پارتو، برآورد δ, α را می توان محاسبه نمود که عبارت اند از: $\hat{\alpha} = 66/9736$ و $\hat{\delta} = 1/0078$. محاسبات همچنین نشان می دهد که ضریب جینی (\hat{g}) برابر با ۰/۰۰۷۵۲، شاخص فقر ($\hat{\alpha}_0$) برابر با ۳۸٪ و احتمال بقاء برابر با ۶۲٪ است. از این نتایج می توان نتیجه گرفت:

الف- توزیع درآمد در نمونه ی مورد مطالعه عادلانه است.

ب- کم تر از ۵۰ درصد افراد نمونه دارای درآمدی کم تر از حد اقل درآمد مورد نیاز برای زندگی هستند.

جدول ۱- فراوانی مشاهدات برحسب درآمدهای 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 ارائه شده بیان می‌کند قصد داریم تا تحت پیشینه‌های در نظر گرفته شده در این پژوهش و بر اساس تابع زیان توان دوم خطاها، برآورد توزیع‌های پسین و برآورد پارامتر مورد نیاز را به دست آوریم. جدول ۲ وضعیت این برآوردگرها را تحت پیشینه‌های مختلف معرفی می‌کند. با توجه به برآوردهای بدست آمده برای پارامترهای اولیه، توزیع‌های پسین به شکل زیر معرفی می‌شود:

تحت نمایی:

$$\pi(\alpha, \delta | x) = \frac{\alpha^{255}}{0/986432} \delta^{255\alpha - 0/5} e^{-\alpha(6/596)} I(\delta) I(\alpha)_{(0,0/718) (0/74,\infty)}$$

تحت گاما:

$$\pi(\alpha, \delta | x) = \frac{\alpha^{256}}{0/995225} \delta^{255\alpha - 0/9} e^{6/296\alpha} I(\delta)_{(0,0/716)}$$

جدول ۲- تعیین برآورد پارامترهای به روش نظریه‌ی بیز

توزیع پیشین α	برآورد پارامترهای اولیه	برآورد پارامتر تحت تابع زیان توان دوم خطاها
نمایی دوپارامتری	$\hat{\lambda} = 0/8$	$\hat{\alpha}_{\beta} = 2/79992$
	$\hat{\mu} = 0/74$	$\hat{\delta}_{\beta} = 0/716993$
	$\hat{\delta}_{\circ} = 0/718$	$\hat{g}_{\beta} = 0/218672$
	$\hat{\beta} = 0/5$	$\hat{\mu}_{\beta} = 1/11918$
		$\hat{\alpha}_{\beta} = 0/6215$
		$\hat{\lambda}_{\beta} = 0/3784$
گاما	$\hat{\alpha}_{\circ} = 2$	$\hat{\alpha}_{\beta} = 2/79826$
	$\hat{b}_{\circ} = 0/5$	$\hat{\delta}_{\beta} = 0/714994$
	$\hat{\delta}_{\circ} = 0/716$	$\hat{g}_{\beta} = 0/156679$
	$\hat{\beta} = 0/1$	$\hat{\mu}_{\beta} = 1/11642$
		$\hat{\alpha}_{\beta} = 0/62424$
	$\hat{\lambda}_{\beta} = 0/37575$	

ملاک مقایسه:

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

جدول ۳- برآورد واریانس برآوردگرها

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

جدول ۴- جدول تحلیل واریانس مربوط به تأثیر عوامل کمکی بر درآمد 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 درآمد خانوارهای نمونه

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

نتیجه‌گیری

می‌توان گفت در این مطالعه روش بیزی پیشنهاد شده بسیار کارا تر از روش فراوانی‌گرا عمل نموده است. البته می‌توان 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 علامه طباطبائی، گروه آمار.
رایانشانی: ffeskandari@yahoo.com

رقیه قبادی

فوق لیسانس آمار

تهران، وزارت بازرگانی، اداره‌ی آمار و برنامه‌ریزی.