روش بیزی برای تجزیه‌وتحلیل ریسک فاجعه‌آمیز زلزله

احسان جعفری[1]

دکتر امیر تیمور پاینده نجف‌آبادی[2]

چکیده

نظریه مقادیر کرانگین به‌طور گسترده‌ای در تجزیه‌وتحلیل ریسک‌های فاجعه‌آمیز کاربرد دارد. برای برآورد توزیع حدی ارزش مازاد در یک سطح آستانه خاص می‌توان از توزیع پارتو تعمیم‌یافته استفاده کرد. بنابراین تجزیه‌وتحلیل رفتارهای دمی بر روی ریسک‌های فاجعه‌آمیز تأثیرگذار می‌باشد. برآورد پارامترهای مدل در GPD و ارزیابی حق‌بیمه ریسک‌های فاجعه‌آمیز به رفتار خسارت‌ها بستگی دارد. در این مقاله از چهار مدل آمیخته برای مدل‌بندی زیان فاجعه‌آمیز زلزله استفاده شده است، همچنین برای برآورد پارامترهای ناشناخته و سطح آستانه مناسب در این مدل‌ها از رویکرد بیزی استفاده می‌کنیم. برای محاسبه برآوردهای بیزی پارامترهای مدل از روش‌های مونت‌کارلوی زنجیره مارکفی و برای مقایسه مدل از ارزش‌های اطلاعات انحراف معیار استفاده می‌کنیم. نتایج نشان می‌دهد که برآورد سطح آستانه و پارامترهای شکل و مقیاس توزیع GPD کاملاً متفاوت است. در انتها، ارزش در معرض خطر برای مدل‌های آمیخته پیشنهادی در سطح‌های اطمینان مختلف محاسبه می‌شود. از نرم‌افزارهای Winbugs و R برای برآورد پارامترها و مناسب‌بودن مدل‌ها استفاده شده است.

واژگان کلیدی: مدل‌های آمیخته، رویکرد بیزی، روش‌های مونت‌کارلوی زنجیر مارکفی، زیان‌های زلزله

---

۱. دانشجوی کارشناسی ارشد آمار بیمه، دانشگاه شهید بهشتی — Ejafari71.n@gmail.com

۲. استاد، دانشکدهٔ علوم ریاضی، دانشگاه شهید بهشتی — Amirtpayandeh@gmail.com

۱- مقدمه

آمار خسارت‌های مالی و جانی زلزله‌های گذشته در کشور و اینکه ایران با قرار گرفتن در کمربند جهانی زلزله و وجود نقاط متراکم جمعیتی در بافت‌های فرسوده، منطقه‌ای با پتانسیل بالا برای ریسک‌های خسارت ناشی از زلزله محسوب می‌شود. زلزله جدی‌ترین تهدید به جان و مال مردم است. گزارش‌ها نشان می‌دهد که سه تا از ده زلزله مرگبار جهان در چین رخ داده است. زلزله سال 1556 در شانشی به‌عنوان مرگبارترین زلزله در تاریخ و بنا به گزارش 80000 نفر کشته شده است. اندازه‌گیری ریسک‌های فاجعه‌آمیز برای شرکت‌های بیمه بسیار مهم است زیرا باعث ورشکستگی شرکت بیمه می‌شود. یکی از روش‌های اندازه‌گیری ریسک در عمل ارزش در معرض خطر(VAR) است. ارزش در معرض خطر یک معیار اندازه‌گیری مخاطره است که حداکثر زیان را با سطح P برآورد می‌کند. ارزش در معرض خطر برای تعیین حق‌بیمه و سرمایه ذخایر، مجموعه‌ای از کسورات بیمه و حق‌بیمه‌ی اتکایی، برآورد ادعا و زیان مورد انتظار محاسبه می‌شود. به تازگی در کمیته نظارتی بازل در سال 2012 به جای VAR از ریسک مورد انتظار ES در مدل‌های داخلی بانک‌ها برای تعیین الزامات سرمایه استفاده شده است. ES در سطح $P \in (0,1)$ که به صورت زیر نشان داده می‌شود:

$$ES[X,P] = E[(X - VAR[X,P])_+]$$

که X نشان‌دهنده یک دارای یا ادعا و $(.)_+$ نشان‌دهنده بخش مثبت یک تابع است.

یک مسأله کلیدی برای برآورد ارزش در معرض خطر انتخاب توزیع مناسب برای X است. معیار اندازه گیری VaR پایه‌ای برای توزیع زیان‌های بزرگ به جای توزیع همه زیان‌ها است و توسط (بال۲۰۰۳، ۲۰۰۸، لانگین ۲۰۰۰ و مک نیل و فری ۲۰۰۰) استفاده شده است. بنابراین ساختار توزیع‌های زیان می‌تواند تأثیر قابل توجهی در برآورد حق‌بیمه داشته باشد. کشورهای در معرض خطر سوانح طبیعی از جمله ایران و چین، پس از وقوع زلزله با مسئله‌ی جبران خسارت ناشی از آن روبه رو هستند. با توجه به اینکه خسارات ناشی از زلزله در چند دهه اخیر افزایش داشته می‌توان با ایجاد انواع پوشش‌های بیمه ای خسارت‌های ناشی از زلزله را جبران نمود.

۲- بیان مسئله

اگر شرکت‌های بیمه حداکثر زیان حوادث فاجعه‌آمیز را محاسبه نکنند امکان ورشکستگی آن‌ها وجود دارد. بنابراین یک سیستم بیمه‌ای موفق و کارامد باید توان محاسبه حداکثر زیان در بیمه زلزله و دیگر حوادث فاجعه‌آمیز را داشته باشد. برای داده‌های بیمه‌ای معمولا از توزیع‌های زیان گاما، لگ نرمال و وایبل استفاده می‌کنند. با توجه به اینکه توزیع‌های زیان اغلب شدیداً دم سنگین هستند این توزیع‌ها برای تجزیه‌وتحلیل زیان‌های حوادث فاجعه‌آمیز به خصوص زلزله مناسب نیستند. نظریه مقادیر کرانگین توسط فیشر و تیپت در سال ۱۹۲۸ ارائه شده و به طور گسترده‌ای در تجزیه‌وتحلیل ریسک‌های فاجعه‌آمیز در امور مالی و بیمه در چند دهه اخیر (کولز۲۰۰۱، اسمیت۲۰۰۳) استفاده شده است. متغیرهای زیان $X_1, \dots, X_n$ دارای توزیع یکسان و مستقل با تابع توزیع ناشناخته H(X) است. $Z_n = \max(X_1, \dots, X_n)$ ماکزیمم متغیرهای زیان است. فیشر و تیپت در سال ۱۹۲۸ تابع توزیع تجمعی از Z_n که همگرا به مقادیر بزرگ تعمیم یافته، تابع زیرا را ذکر کردن

$$G(Z) = \begin{cases} \exp\left(-\left(1+\varepsilon\dfrac{z-\mu}{\sigma}\right)^{-\frac{1}{\varepsilon}}\right) & \varepsilon \neq 0\,,\ \ 1+\varepsilon z > 0 \\ \exp\left(-exp\left(-\dfrac{z-\mu}{\sigma}\right)\right) & \varepsilon = 0\,, -\infty < \ z < \infty \end{cases} \quad (1)$$

که در آن μ، σ و ε پارامترهای مکانی، مقیاسی و شکل است. این تابع یک روش برآورد ماکسیم فراهم می‌کند. با این حال اگر داده‌های دیگر در مقادیر بزرگ بالاتر از آستانه باشند مدلسازی تمام ماکسیم بلوک‌ها بی فایده است. یک روش موثر، پیک بالای سطح آستانه برای مدلسازی رفتاری از ارزش‌های بزرگ بیشتر از سطح آستانه باشد استفاده می‌شود. پیک بالای سطح آستانه بر روی مدلسازی که عبور کند از یک متغیر تصادفی $\boldsymbol{x}$ بیش از یک آستانه بالا ثابت، $\boldsymbol{u}$ تمرکز دارد. $\boldsymbol{X \sim F}$ نشان‌دهنده مقدار مورد انتظار (به‌عنوان مثال، ارزش زیان یا بازده دارایی)، $\boldsymbol{Y = X - U}$ نشان‌دهنده بیش از $\boldsymbol{U}$ است. پیکاندوز در سال ۱۹۷۵ و بالکمان وهان در سال ۱۹۷٤ توزیع حدی از $\boldsymbol{Y}$، با توجه به اینکه $\boldsymbol{X > U}$ است نشان می‌دهند، توزیع پارتو تعمیم یافته به صورت زیر می‌باشد:

$$g(y|\varepsilon,\sigma) = \begin{cases} \sigma^{-1}(1+y\frac{\varepsilon}{\sigma})^{-\frac{(1+\varepsilon)}{\varepsilon}} & if\ \left(1+y\frac{\varepsilon}{\sigma}\right) > 0 \\ \sigma^{-1}\exp\left(\frac{-y}{\sigma}\right) & \varepsilon = 0 \end{cases} \tag{2}$$

پارامتر شکل ε و مقیاس $\sigma > 0$ زمانیکه u میل می‌کند به نقطه بالای از دامنه $H(X)$ است. تابع توزیع تجمعی از توزیع پارتو تعمیم یافته به صورت زیر است:

$$G(y|\varepsilon,\sigma) = \begin{cases} 1-\left(1+y\frac{\varepsilon}{\sigma}\right)^{-\frac{1}{\varepsilon}} & if\ \varepsilon \neq 0 \\ 1-\exp\left(\frac{-y}{\sigma}\right) & \varepsilon = 0 \end{cases} \tag{3}$$

مدل تابع توزیع تجمعی پارتو تعمیم یافته شکل ساده و به‌طور گسترده در تجزیه‌وتحلیل مقادیر بزرگ کاربرد دارد. برآورد آستانه U در مدل توزیع تجمعی پارتو تعمیم یافته دشوار است، زیرا تعادل بین اریبی با توجه به تقریب دم مجانبی و عدم قطعیت از برآورد پارامتر ناشی از حد آستانه داده است. اسکارت و مک دونالد در سال ۲۰۱۲ رویکردی (لدباتر در سال ۱۹۸۳) در مدلسازی کلاسیک آستانه ثابت، آستانه را قبل از برازش کردن با استفاده از تشخیص گرافیکی بررسی کردند. این روش ساده است چرا که نیاز به تخصص قابل توجهی در تجزیه‌وتحلیل ندارد. علاوه‌براین داشتن بیش از یک آستانه مناسب امکان‌پذیر است. مقادیر درستنمایی توزیع تجمعی پارتو تعمیم یافته برای آستانه‌های مختلف با تغییر اندازه دشوار می‌باشد. علاوه‌براین، حداکثر برآوردگر درستنمایی شرایط نظم در آن رعایت نشده است اگر $\varepsilon \epsilon (-1,-.5)$ و آن وجود ندارد اگر $\varepsilon < -.5$ باشد (اسمیت ۱۹۸۳). یک مدل آمیخته (برنس۲۰۰٤ و کاریو و بنگو ۲۰۰۹) برای حل این مشکل اخیرا توسعه داده‌اند. آستانه در این مدل آمیخته به‌عنوان پارامتر درمان با استفاده از روش بیزی پیشنهادشده توسط (برنس۲۰۰٤) تخمین زده می‌شود. مدل آمیخته فرض می‌شود که مشاهده‌هایی از متغیر تصادفی X زیر آستانه U از یک توزیع زیان $H(.|\theta_1)$ آمده باشند، درحالی‌که داده‌هایی که بالاتر از آستانه باشند از توزیع دم سنگین $G(x|\theta_1,u)$ آمده‌اند، θ_1 و θ_2 پارامترهای بردار در $H(.)$ و $G(.)$ هستند. تابع توزیع تجمعی از X را می‌توان به صورت زیر نوشت:

$$F(x|\theta_1,\theta_2,u) = \begin{cases} H(x|\theta_1) & x < u \\ H(u|\theta_1) + [1-H(u|\theta_1)]G(x|\theta_2,u) & x \geq u \end{cases} \tag{4}$$

که $\theta = (\theta_1,\theta_2,u)$ بردارهای پارامتر هستند. برای یک نمونه $x = (x_1,\dots,x_n)$ تابع درستنمایی را می‌توان به صورت زیر نوشت:

$$L(\theta;x) = \prod_{\{i;x_i<u\}} h(x_i|\theta_i) \prod_{\{i;x_i\geq u\}} \left(1-H(u|\theta_1)\right)g(x_i|\theta_2,u)$$

که $h(.)$ تابع چگالی از توزیع $H(.)$ است، $g(x_i|\theta_2,u)$ تابع چگالی از توزیع $G(.)$ است. فریگس در سال ۲۰۰۲ از توزیع وایبل برای بخش مرکزی و توزیع پارتو تعمیم یافته برای بخش بزرگ استفاده کرد. برنس در سال ۲۰۰٤ یک آمیخته از توزیع گاما و توزیع پارتو تعمیم یافته برای بخش دم در نظر گرفت. ناسینتو در سال ۲۰۱۲ به کاربرد آمیخته‌ای از گاماها به بخش مرکزی و یک توزیع تجمعی پارتو تعمیم یافته برای بخش دم در نظر گرفت. این مطالعه چهار مدل‌های آمیخته در تجزیه‌وتحلیل ریسک زلزله پیشنهاد می‌کند که عبارتند از توزیع‌های گاما و پارتو، آمیخته گاما و پارتو، لگ نرمال و پارتو تعمیم یافته و ترکیب توزیع وایبل و توزیع پارتو تعمیم یافته، روش بیزی برای برآورد پارامترهای مدل و استفاده از اطلاعات انحراف معیار (DIC) لین در سال ۲۰۰۲ برای ترکیب مدل استفاده کرد. علاوه بر این مطالعه محاسبه VARو ES برای اندازه گیری ریسک است.

بخش ۲،۱ مدل‌های آمیخته متفاوت، ارتباط فرآیند برآورد بیزی و معیارهای مقایسه مدل. بخش ۳ اندازه‌گیری ریسک با VAR و ES برای مدل‌های آمیخته متفاوت است. بخش ٤ آنالیز داده زیان زلزله با برازش این مدل‌های آمیخته با روش بیزی. بخش ٥ بحث و نتیجه گیری است.

۳- مدل‌های آمیخته و استنباط بیزی

مدل آمیخته ٤ دارای بسیاری از ترکیبات ممکن است. چهار ترکیب متفاوت از مدل آمیخته در نظر می‌گیرد، در آن به‌طور منظم توزیع زیان $H(.)$ می‌تواند گاما، گامای آمیخته، لگ نرمال و وایبل در نظر بگیرد، درحالی‌که توزیع تجمعی پارتو تعمیم یافته توزیع شرطی حدی از مازاد است.

۳-۱ توزیع گاما و برآورد بیزی

با پارامتر شکل $\alpha > 0$ و پارامتر نرخ $\beta > 0$، احتمال تابع چگالی از توزیع گاما به صورت زیر است:

$$f_G(x|\alpha,\beta) = \frac{\beta^{\alpha}}{(\alpha-1)!}\exp(-\beta x)\, x^{\alpha-1} \quad x>0 \tag{5}$$

با تابع توزیع تجمعی که به صورت $F_G(x|\alpha,\beta) = \int_0^X f_G(t|\alpha,\beta)dt$ نشان داده می‌شود. مدل آمیخته (٤) به صورت زیر نوشته می‌شود:

$$F_G(x|\alpha,\beta,\varepsilon,\sigma,u) = \begin{cases} F_G(x|\alpha,\beta) & x<u \\ F_G(u|\alpha,\beta) + [1-F_G(u|\alpha,\beta)]G(x|\varepsilon,\sigma,u) & x \ge u \end{cases} \tag{6}$$

که $G(x|\varepsilon,\sigma,u)$ تابع توزیع تجمعی از تابع چگالی (۳) است. تابع درستنمایی مدل m_0 به صورت زیر است:

$$L(\theta;x) \begin{cases} \prod_{\{i;x_i<u\}} f_G(x_i|\alpha,\beta) \prod_{\{i;x_i\ge u\}} [1-F_G(u|\alpha,\beta)]g((x_i-u)|\varepsilon,\sigma) & \varepsilon \ne 0 \\ = \prod_{\{i;x_i<u\}} f_G(x_i|\alpha,\beta) \prod_{\{i;x_i\ge u\}} [1-F_G(u|\alpha,\beta)]g((x_i-u)|\varepsilon,\sigma) & \varepsilon = 0 \end{cases} \tag{7}$$

که $\theta = (\alpha,\beta,\varepsilon,\sigma,u)$ است. یک فرآیند بیزی توسعه از برآورد این مدل است. توزیع‌های پیشین از پارامترهای (ε,σ) در توزیع پارتو تعمیم یافته نیاز دارند. توزیع پیشین توسط کابرس در سال ۲۰۰۷ دنبال می‌کند.

$$\pi(\varepsilon,\sigma) \propto \sigma^{-1}(1+\varepsilon)^{-1}(1+2\varepsilon)^{\frac{-1}{2}} \quad \varepsilon > \frac{-1}{2}\ ,\ \sigma>0 \tag{8}$$

این توزیع پیشین با استفاده از قاعده جفری تعریف شده است. کابرس در سال ۲۰۰۷ اعلام کرد که این توزیع پیشین هدایت می‌یابد به توزیع پسین که مناسب است و استنباط بیزی دقیق‌تر از دیگر توزیع‌های پسین است. توزیع پسین (ε,σ) به صورت زیر است:

$$\pi(\varepsilon,\sigma|x_1,\dots,x_n) \propto \sigma^{-(n+1)}(1+\varepsilon)^{-1}(1+2\varepsilon)^{\frac{-1}{2}}\prod_{i=1}^{n}\left(1+x_i\frac{\varepsilon}{\sigma}\right)^{-\frac{(1+\varepsilon)}{\varepsilon}} \qquad (9)$$

$$\varepsilon > \frac{-1}{2} \quad , \quad \frac{\varepsilon}{\sigma} > -\frac{1}{\max\{x_1,\dots,x_n\}} \ ,\sigma > 0$$

برای آستانه u، توزیع نرمال $N(\mu_n,{\sigma^2}_n)$ به‌عنوان توزیع پیشین گرفته شده است (برنس۲۰۰٤). سیمنتو و همکاران در سال ۲۰۱۲ اشاره کردند که میانگین μ_n باید بزرگتر از نمونه باشد، پیشین نباید بیش از حد متمرکز شود مگر اینکه دانش قبلی قابل توجهی در مورد این پارامتر در بر داشته باشد. توزیع پیشین برای پارامترهای (ε,σ) پیدا شده است و u در مدل‌های آمیخته می‌باشد. برای پارامترهای (α,β)، توزیع‌های پیشین رایج استفاده می‌شود:

$$\alpha \sim IG(a,b)\ , \qquad \beta \sim Gamma\left(\frac{c^2}{d},\frac{c}{d}\right) \qquad (10)$$

با پارامترهای مثبت a، b، c، d که IG را توزیع معکوس گاما نشان می‌دهد. تحت پیشین در (۸) و (۱۰) و نمونه $x=(x_1,\dots x_n)$ و چگالی پسین از θ را می‌توان به صورت زیر نوشت:

$$\pi_0(\theta|x) \propto \prod_{\{i;x_i<u\}}\frac{\beta^\alpha}{(\alpha-1)!}\exp(-\beta x_i)\,{x_i}^{\alpha-1}\prod_{x_i\ge u}\left\{[1-F_G(u|\alpha,\beta)]\left[\frac{1}{\sigma}\left(1+\varepsilon\frac{x_i-u}{\sigma}\right)\right]\right\} \qquad (11)$$
$$\frac{b^\alpha}{(a-1)!}\exp\left(-\frac{b}{a}\right)\alpha^{-a-1}\frac{d^c}{(c-1)!}\exp(-\beta d)\,\beta^{c-1}\sigma^{-1}(1+\varepsilon)^{-1}(1+2\varepsilon)^{-\frac{1}{2}}$$
$$\frac{1}{\sigma\sqrt{2\pi}}\ \exp\{-\frac{(u-\mu_u)^2}{2{\sigma^2}_u}\}$$

فرم چگالی پسین در معادله ۱۱ درنمونه پیچیده و دشوار می‌باشد. بنابراین الگوریتم متروپلیس هستینگس و نمونه گیری گیبز برای بدست آوردن برآورد بیز از این چگالی پسین استفاده می‌شود.

۳-۲- توزیع گاما آمیخته و برآورد بیز نیمه پارامتری

مدل‌سازی آمیخته به‌طور گسترده در تجزیه و تحلیل داده‌های پیچیده استفاده می‌شود. چگالی ناپارامتری را می‌توان با استفاده از توزیع نرمال آمیخته برآورد کرد. چگالی از داده زیان می‌توان با یک توزیع آمیخته گاما مناسب باشد. آمیخته متناهی K تایی از توزیع گاما[1] می‌توان بیان شود:

$$f_{MG}(x|\alpha,\beta,p) \quad = \sum_{j=1}^{k} p_j f_G\left(x|\alpha_j,\beta_j\right), \qquad (12)$$

که $f_G\left(x|\alpha_j,\beta_j\right)$ در معادله ٥ می‌باشد. $\alpha=(\alpha_1,\dots,\alpha_k)$و $\beta=(\beta_1,\dots,\beta_k)$ و $p=(p_1,\dots,p_k)$ و $\sum_{j=1}^{k}p_j$ و $\gamma=(\alpha,\beta,p)$ می‌باشد. مدل آمیخته ٤ را می‌توان به صورت زیر نوشت:

$$F(x|\gamma,\varepsilon,\sigma,u) = \begin{cases} F_{MG}(x|\gamma) & x<u \\ F_{MG}(u|\gamma)+[1-F_{MG}(u|\gamma)]G(x|\varepsilon,\sigma) & x\ge u \end{cases} \qquad (13)$$

1. MG

که $F_{MG}(x|\gamma)=\sum_{j=1}^{k}p_j\int_0^x f_G(u|\alpha_j,\beta_j)du$ این مدل را با m_1 نشان می‌دهند و تابع درستنمایی از m_1 به‌صورت زیر می‌باشد:

$$L(\theta;x)=\begin{cases}\prod_{\{i;x_i<u\}}f_{MG}(x_i|\gamma)\ \prod_{\{i;x_i\geq u\}}\left(1-F_{MG}(u|\gamma)\right)g\left((x_i-u)|\varepsilon,\sigma\right), & \varepsilon\neq 0\\ \prod_{\{i;x_i<u\}}f_{MG}(x_i|\gamma)\ \prod_{\{i;x_i\geq u\}}\left(1-F_{MG}(u|\gamma)\right)g\left((x_i-u)|\varepsilon,\sigma\right), & \varepsilon=0\end{cases}\quad(14)$$

که $\theta=(\gamma,\varepsilon,\sigma,u)$ است.

رویکرد بیزی نیمه‌پارامتری توسعه یافته به منظور برآورد پارامترها در معادله ۱۴ می‌باشد. برخی از محدودیت‌ها باید در پارامترهای مدل تحمیل شود زیرا مدل آمیخته فاقد شناسایی هستند. توزیع آمیخته گاما به شرح زیر داده شده است:

$c(\alpha)=\{\alpha|0<\alpha_1<\cdots.<\alpha_k\}$

برای پیاده سازی روش بیزی دنبال کنید پیشین در نظر گرفته برای α

$$p(\alpha_1,\dots,\alpha_k)=K\prod_{i=1}^{k}f_{IG}(\alpha_i|a_i,b_i)I(\alpha_1<\cdots<\alpha_k)\quad(13)$$

که $K^{-1}=\int_{C(\alpha)}\prod_{i=1}^{k}f_{IG}(\alpha_i|a_i,b_i)d(\alpha_1,\dots,\alpha_k)$ و $f_{IG}(.)$ تابع چگالی وارون گاما را نشان می‌دهد.

۳-۳- توزیع لگ نرمال و برآورد بیزی

یک متغیر تصادفی مثبت X با توزیع لگ نرمال تابع چگالی زیر را دنبال می‌کند:

$$f_{LN}(x|\mu,\sigma_l)=\frac{1}{x\sigma_l\sqrt{2\pi}}\exp\left\{-\frac{(ln(x)-\mu)^2}{2\sigma^2{}_l}\right\},x>0\quad(17)$$

با پارامتر مکانی μ، پارامتر مقیاس σ_l و تابع توزیع تجمعی

$$F_{LN}(x|\mu,\sigma_l)=\phi\left(\frac{\ln x-\ \mu}{\sigma_l}\right),x>0,\quad(18)$$

که ϕ، CDF توزیع نرمال است. اگر متغیر تصادفی زیان X زیر استانه U باشد فرض می‌شود توزیع لگ نرمال سپس مدل آمیخته ٤ می‌توان به صورت زیر نوشت:

$$F(x|\mu,\sigma_l,\varepsilon,\sigma,u)=\begin{cases}F_{LN}(x|\mu,\sigma_l), & x<u\\ F_{LN}(u|\mu,\sigma_l)+[1-F_l(u|\mu,\sigma_l)]G(x|\varepsilon,\sigma,u) & x\geq u\end{cases}\quad(19)$$

$\theta=(\mu,\sigma_l,\varepsilon,\sigma,u)$. این مدل M2 نشان می‌دهد. وارون مشاهده نمونه x ، تابع درستنمایی M2 می‌توان نوشت:

$$L(\theta;x)=\begin{cases}\prod_{\{i;x_i\leq u\}}f_{LN}(x_i|\mu,\sigma_l)\prod_{\{i;x_i\geq u\}}\left(1-F_{LN}(u|\mu,\sigma_l)\right)\left(\frac{1}{\sigma}\left[1+\varepsilon\frac{x_i-u}{\sigma}\right]^{-\frac{(1+\varepsilon)}{\varepsilon}}_{+}\right), & \varepsilon\neq 0\\ \prod_{\{i;x_i\leq u\}}f_{LN}(x_i|\mu,\sigma_l)\prod_{\{i;x_i\geq u\}}\left(1-F_{LN}(u|\mu,\sigma_l)\right)\left(\frac{1}{\sigma}\left[1+\varepsilon\frac{x_i-u}{\sigma}\right]^{-\frac{(1+\varepsilon)}{\varepsilon}}_{+}\right), & \varepsilon\neq 0\end{cases}\quad(20)$$

برای بدست آوردن برآورد بیز از مدل M2، دنبال می‌کنیم پیشین منظم و دقیق برای پارامترها

$$\mu\sim N\left(\mu_{\mu0},{\sigma_{\mu0}}^2\right),\pi({\sigma_l}^2)\propto({\sigma_l}^2)^{-1}\quad(21)$$

که $\mu_{\mu0}$ و ${\sigma_{\mu0}}^2$ پارامتر می‌باشند. توزیع پسین از θ در لگاریتم مقیاس می‌تواند بدست آورد:

$$\log\pi_2(\theta|x)=c_2+\sum_{i;x_i\leq u}[-\ln(x_i)-\ln(\sigma_l)-\frac{(\ln(x_i)-\mu)^2}{2{\sigma_l}^2}$$

$$+\sum_{x_i>u}\{\ln[1-F_{LN}(u|\mu,\sigma_l^2)]-\ln(\sigma)+\ln\left(1+\varepsilon\frac{x_i-u}{\sigma}\right)\}$$
$$-\ln(\sigma_l^2)-\frac{\left(u-\mu_{\mu0}\right)^2}{2{\sigma_{\mu0}}^2}-\ln(\sigma_l^2)$$
$$-\ln(\sigma)-\ln(1+\varepsilon)-\frac{1}{2}\ln(1+2\varepsilon)-\ln(\sigma_u^2)-\frac{(u-\mu_u)^2}{2{\sigma_u}^2} \qquad (22)$$

لگاریتم نرمالایز c_2 ثابت است. یک توزیع لگ نرمال آمیخته می‌توان برای برازش چگالی از داده زیان در نظر گرفت. در مدل M3 می باشد.

۳-٤- توزیع وایبل و برآورد بیزی

پارامتر مقیاس λ و پارامتر شکل η است. تابع چگالی $f_W(x|\lambda,\eta)$ از تابع توزیع وایبل $F_W(x|\lambda,\eta)$ به صورت زیر است:

$$f_W\left(x|\lambda,\eta\right)=\frac{\eta}{\lambda}(\frac{x}{\lambda})^{\eta-1}\exp\left\{-\left(\frac{x}{\lambda}\right)^{\eta}\right\},\ \ x\geq 0 \qquad (23)$$
$$F_W\left(x|\lambda,\eta\right)=1-\exp\left\{-\left(\frac{x}{\lambda}\right)^{\eta}\right\},x\geq 0 \qquad (24)$$
$$E(x)=\lambda\Gamma\left(1+\frac{1}{\eta}\right),Var(x)=\lambda^2[\lambda\Gamma\left(1+\frac{2}{\eta}\right)-(\Gamma\left(1+\frac{1}{\eta}\right))^2]$$

متغیر زیان X با آستانه u توزیع وایبل دنبال می‌کند، مدل آمیخته(٤) به صورت زیر نوشته می‌شود:

تابع درستنمایی می‌توان به صورت زیر نوشت:

$$L(\theta;x)$$
$$=\begin{cases}\prod_{\{i;\,x_i<u\}}f_W(x_i|\theta_1)\prod_{\{i;x_i\geq u\}}\left(1-F_W(u|\theta_1)\right)(\frac{1}{\sigma}\left[1+\varepsilon\frac{x_i-u}{\sigma}\right]_+^{-\frac{(1+\varepsilon)}{\varepsilon}} & \varepsilon\neq 0\\ \prod_{\{i;\,x_i<u\}}f_W(x_i|\theta_1)\prod_{\{i;x_i\geq u\}}\left(1-F_W(u|\theta_1)\right)(\frac{1}{\sigma}\left[1+\varepsilon\frac{x_i-u}{\sigma}\right]_+^{-\frac{(1+\varepsilon)}{\varepsilon}} & \varepsilon=0\end{cases} \qquad (26)$$

این مدل را با m_3 نشان می‌دهند. توزیع پیشین برای پارامتر وایبل می‌توان به صورت زیر نوشت:

$$\pi\left(\lambda,\eta\right)\propto\left(\frac{1}{\lambda\eta}\right)^{2c} \qquad (27)$$

که $c>0$ است. پسین در لگاریتم مقیاس به صورت زیر است:

$$\log\pi_3(\theta|x)=\ c_3+\sum_{i;\,x_i\leq u}[\ln\eta+\ln\lambda+(\eta-1)(\ln x_i-\ln\lambda)-\left(\frac{x_i}{\lambda}\right)^{\eta}]$$
$$+\sum_{i;x_i>u}\{-\left(\frac{u}{\lambda}\right)^{\eta}-\ln\sigma+\ln(1+\varepsilon\frac{x_i-u}{\sigma})\}$$
$$-2c\left(\ln\lambda+\ln\eta\right)-\ln\sigma-\ln(1+\varepsilon)-\frac{1}{2}\ln(1+2\varepsilon)-\ln{\sigma^2}_u-\frac{(u-\mu_u)^2}{2{\sigma^2}_u} \qquad (28)$$

با ثابت لگاریتم نرمالایز c_3 است.

۳-۵- مقایسه مدل بیزی

در زمینه استنباط بیزی، بسیاری از مقایسه مدل‌ها مانند عامل بیز (کس۱۹۹۵)، آمار براساس معیار (گلفند ۱۹۹۸) و انحراف اطلاع معیار[1] ارائه شده است. استفاده از عوامل بیز در آمار یک جایگزین بیز برای آزمون فرضیه کلاسیک است. محاسبات عوامل بیز دشوار است، اگرچه روش‌های مختلفی برای محاسبه آن ارائه شده است. علاوه‌براین، مقایسه مدل با عامل بیز به توزیع پیشین از پارامترها حساس است. DIC به‌طور گسترده‌ای در انتخاب مدل بیزی که اطلاعات آکائیک می‌گویند استفاده می‌شود. برای مدل‌های رقیب m_k باتوجه به بردار پارامتر مجهول θ_k با بعد d_k، $\{{\theta_k}^s, s = 1, \dots, s\}$ یک نمونه از توزیع پیشین آن شبیه سازی شده است. DIC برای مدل m_k تعریف می‌شود

$$DIC_k = -\frac{2}{J}\sum_{s=1}^{S} \log p\ (x|{\theta_k}^{(s)}, m_k) + 2d_k$$

برای مقایسه مدل، مدل با کوچکترین DIC به‌عنوان مدل نهایی انتخاب می‌شود. برای خانواده نمایی مناسب است. بنابراین DIC به‌عنوان مدل استفاده روش انتخاب شده در این مقاله است.

٤- اندازه‌گیری ریسک براساس یک مدل آمیخته

برنامه‌های کاربردی مالی و بیمه، VAR حداکثر زیان که با احتمال داده شده یا سطح اطمینان عبور نمی‌کند، به مقدار یک نزدیک است اما بیشتر از یک نیست، در طی یک دوره معین از زمان، کوین در سال ۲۰۰٦ اشاره کرد که ارزش در معرض خطر جذاب است .

به‌عنوان مثال: ارزش در معرض خطر هرگونه اطلاعاتی در مورد شدت زیان نمی‌دهد. برخ از اقدامات دیگر از ریسک باید در نظر گرفته شود مانند ES و دم ارزش در معرض خطر $TVAR$ است. برای یک سطح اطمینان $p \in (0,1)$ و ESاست به صورت $ES[x;p] = E[\left(X - VAR(X;p)\right)_+]$

ES را می‌توان به برآورد حق‌بیمه خالص زیان در موقعیت فرضی که بیش از حد به $d = VAR[x;p]$ نیازمند است. $TVAR$ برخلاف VAR است. به طور کامل فراتر از صدک مشخص به جای یک نقطه که می‌تواند به طور گسترده مشاهده بدترین نتایج است.

$$TVAR[x;p] = \frac{1}{1-p}\int_p^1 VAR[x;t]dt$$

TVAR می‌تواند برای اندازه گیری سرمایه متوسط استفاده شود. بنابراین ES، VAR و TVAR باید برای تحلیل ریسک در نظر گرفته شود.

برای مدل آمیخته (٤)، VaR به صورت زیر محاسبه می‌شود

$$VaR[x;p] = \begin{cases} H^{-1}(p), & if\ H(u) > p; \\ u + \dfrac{((1-p)^{-\varepsilon} - 1)\sigma}{\varepsilon}, & if\ H(u) < p \end{cases}$$

با داشتن VaR می‌توانیم به‌دست آوریم.

$$ES(x;p) = E[(x - VaR[x;p])\ |x >\ VaR[x;p]] = \frac{\sigma +\ \varepsilon(VaR[x;p])}{1-\varepsilon}$$

1. Deviance Information Criterion

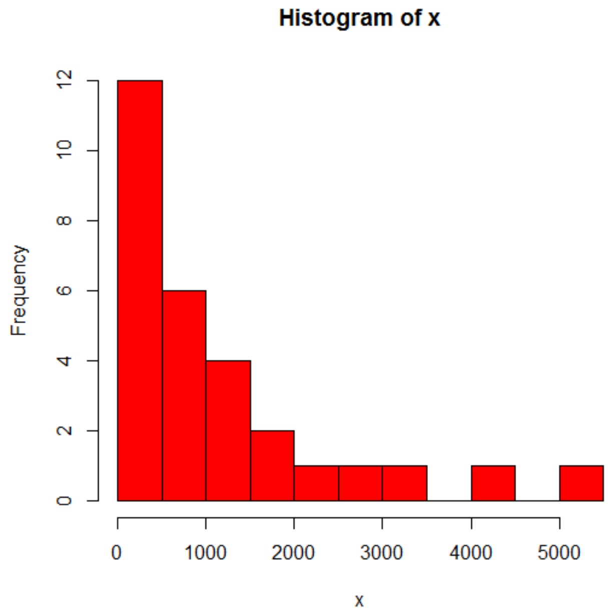
و $TVaR$ می‌توان به صورت زیر نوشت

$$TVaR(x;p) = E\big[x\big|x > VaR[x;p]\big] = VaR[x;p] + ES(x;p).$$

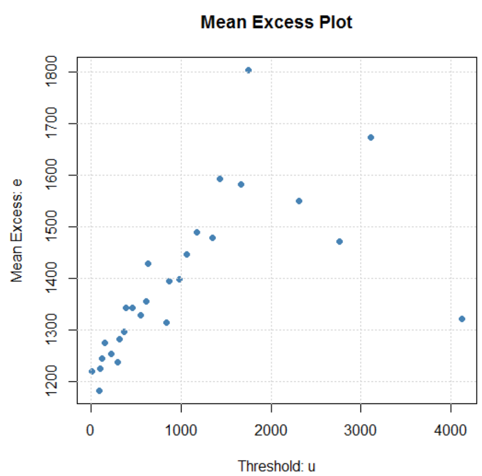
۵- نتایج عددی

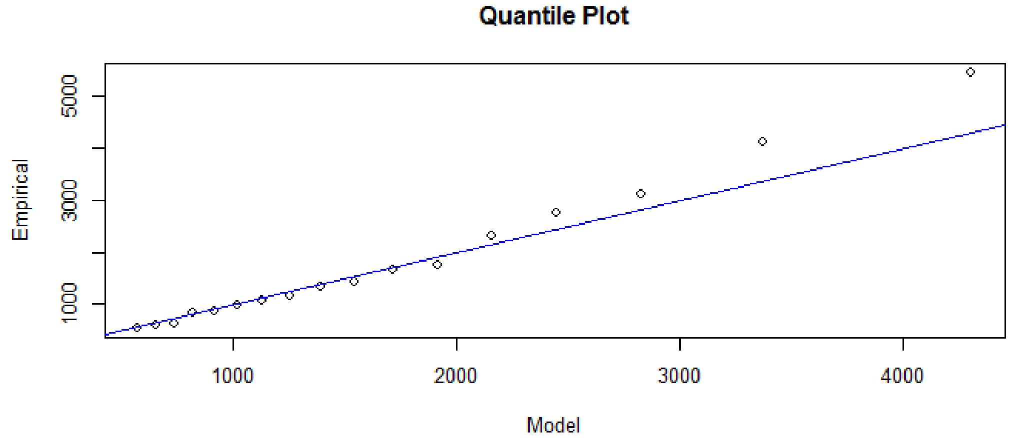
در این مقاله به ریسک زلزله در استان یونان مورد تجزیه و تحلیل قرار می‌گیرد. درباره ۱۱٤ زیان‌های اقتصادی تنظیم تولید ناخالصی داخلی که توسط زلزله ایجاد می‌شود جمع‌آوری شده است.

هیستوگرام در شکل زیر رنج گسترده از داده را نشان می‌دهد. این شکل آمیخته‌ای از توزیع‌های مختلف زیان وجود دارد. دم احتمال یک رفتار دم سنگین نشان می‌دهد. توزیع گاما یا آمیخته توزیع گاما ممکن است بیشتر برای بخش مرکزی مربوطه است. همان‌طورکه از شکل زیر معلوم است داده‌ها تا ۲۰۰۰ به خوبی برازش شده‌اند اما از آن به بعد بیش برآوردی دارد و نشان می‌دهد که داده‌ها خوب برازش نشده است.



QQ plot و میانگین مازاد نمونه در شکل زیر رسم شده است که یک رفتار دم سنگین دارد. رسم QQ plot یک راهنمایی مناسب برای دم سنگین است و می‌تواند استفاده قرار گیرد آیا ضرر و زیان از توزیع نمایی آمده است. QQ plot نشان می‌دهد که دم داده سنگین‌تر از توزیع نمایی می‌باشد.





علاوه بر این تابع میانگین مازاد تجربی می‌توان درنظر گرفت. توزیع نمونه $x_1, \dots, x_n$ مستقل و یکسان از $F(x)$ است. تابع تجربی ME نوشته می‌شود به صورت زیر:

$$e(u) = E[x - u | x > u]$$

با برآورد

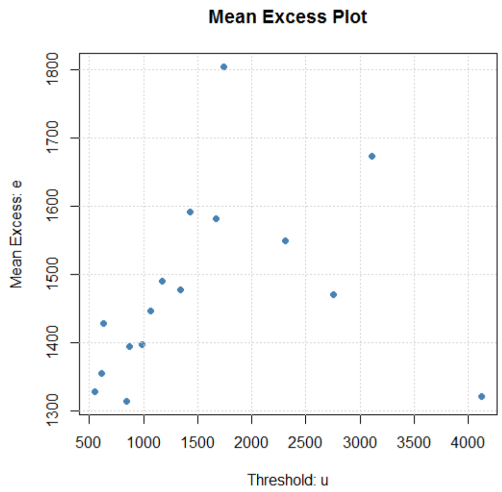
$$e_n(u) = \frac{\sum_{i=1}^{n} (x_i - u)^+}{\sum_{i=1}^{n} 1_{x_{i>u}}}$$

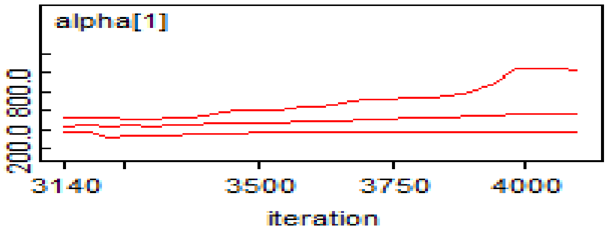
که $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ است.

نمودار ME را می‌توان با نقاط $\{(u, e_n(u)), x_1 < u < x_n\}$ برای بررسی اینکه آیا داده‌ها از یک مدل توزیع GPD آمده‌اند بررسی کنیم. اگر طرح ME نزدیک به خط برازش بالا از آستانه باشد، هیچ شواهدی خلاف استفاده از مدل توزیع GPD ارائه نشده است (دیوچون ۱۹۹۰). شکل دو میانگین نمونه طرح بیش از حد را نشان می‌دهد. طرح دارای یک روند رو به بالا و نزدیک به خطی است که ضرر و زیان بزرگتر از آستانه است. این نشانه‌ای از رفتارهای دم سنگین، یک مدل توزیع پارتو تعمیم یافته مورد را به تناسب ارزش‌های بزرگ از این مجموعه مورد استفاده قرار می‌گیرد از این مجموع داده.

۵–۱– مدل

داده زیان آنالیز می‌شود با چهار مدل $m_0\ m_1\ m_2 m_3$ در نظر می‌گیریم. روش مونت کارلو زنجیر مارکفی برای بدست آوردن برآورد بیز پارامترهای مدل مورد استفاده قرار می‌گیرد. سه زنجیر با اجرا حروف اول مختلف برای بررسی همگرایی الگوریتم است. اثری از پارامترها برای چهار مدل در شکل سه است. ارائه زنجیر پس از ٤٠٠٠ تکرار برای آمیخته همه مدل است. بنابراین استنباط بیزی پس از ٤٠٠٠ تکرا بدست آمده. برآورد بیزی، انحراف استاندارد، فاصله اطمینان ۹۵ درصد از همه پارامترها را برای چهار مدل در جدول دو است. برازش برآورد آستانه در مدل m_0 کاملا مشابه m_2 ، m_3 اما بسیار متفاوت از مدل m_1 است. ارزش DIC نیز در این جدول گزارش به مقایسه عملکرد مدل‌های مختلف است. m_2 کوچکترین مقدار DIC و سایر مدل‌ها دارای ارزش DIC مشابه است. بنابراین عملکرد مدل $m_0\ m_1\ m_3$ قابل مقایسه است. m_2 بهترین عملکرد را دارد..





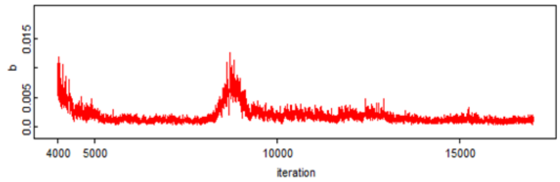
برآورد مدل نشان داد که ES، VAR و TVAR تحت سه فاصله اطمینان ۹۰ درصد، ۹۵ درصد و ۹۹ درصد محاسبه می‌شود. نتایج به‌دست آمده در جدول سه گزارش شده است. این نتایج می‌تواند به شرکت‌های بیمه تصمیم گیری مدیریت ریسک مانند پوشش فاجعه‌آمیز، سطح بیمه اتکایی یا ذخایر فاجعه آمیز کمک می‌کند. برای مثال مدل m_2 انتخاب شده است $VAR[x;90\%]=13.89$ که پوشش بیمه باید ۱۳۸۹۰۰۰۰۰۰ یوان برای پوشش سطح ۹۰درصد زیان ناشی از زلزله $ES[x;90\%]=147.31$ بدان معنی است که بیش از حد مورد انتظار زیان فراتر 13.89 مقدار 147310000000 یوان، می‌توان آن را مورد استفاده برای محاسبه حق‌بیمه اتکایی که باتوجه به بیش از حد بیمه اتکایی زیان، $TVAR[x;90\%]=161.20$ نشان می‌دهد زیان متوسط بالا 13.89 میلیارد 161.20 میلیارد دلار است. این اقدام می‌تواند برای تخمین ذخایر فاجعه‌آمیز شرکت بیمه مناسب باشد.

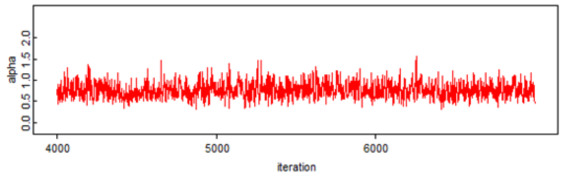
٦- نتیجه

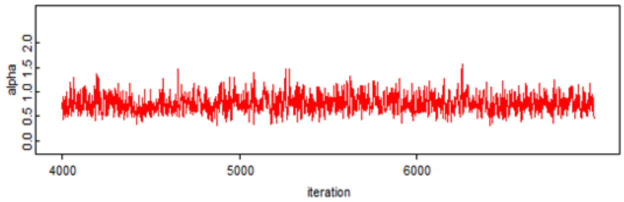
چهار مدل آمیخته برای تجزیه و تحلیل زیان زلزله پیشنهاد شده است، از روش بیزی برای برآورد پارامترهای مدل آمیخته و معیار DIC برای مقایسه استفاده شده است و عملکرد مدل‌های مختلف نمایش داده شده است. نتایج مربوط نشان می‌دهد که مدل آمیخته پیشنهادشده برای داده‌های زیان زلزله مناسب است. پس از به‌دست آوردن برآورد مدل سپس مقدار ریسک با ES، VAR و TVAR محاسبه می‌شود. اندازه‌گیری ریسک‌ها، امنیتی برای شرکت‌های بیمه‌ای به وجود می‌آورد که در تصمیم‌گیری‌های آن‌ها ازجمله، قیمت‌گذاری قراردادهای فردی، مدیریت بیمه، شرکت بیمه اتکایی و تنظیمات کلی در صنعت بیمه کمک می‌کند زیرا یک مسئله مهم برای نرخ‌گذاری بیمه زلزله، اندازه‌گیری عوامل ریسک و توسعه نرخ بیمه است.

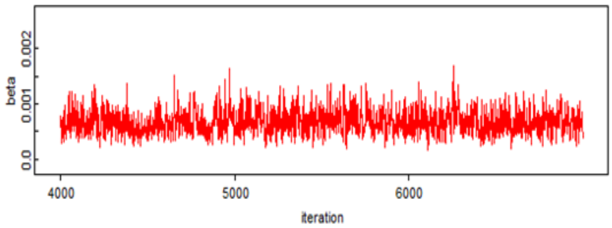
پیوست

مدل گاما









پارامترهای برآوردشده مدل گاما

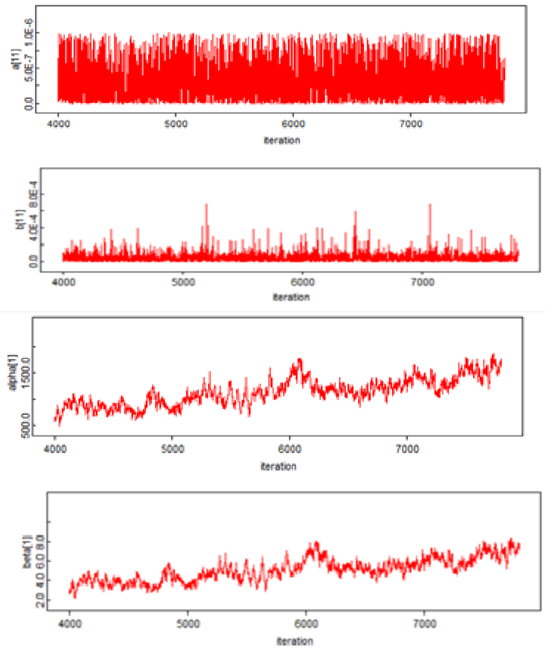
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
a	2.52	4.42	0.3972	0.7785	1.454	15.23	3001	14000

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
b	0.001818	0.001431	1.212E-4	5.87E-4	0.001387	0.006408	4001	13000

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
alpha	0.763	0.176	0.006495	0.4628	0.7449	1.145	4001	3000

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta[1]	1.343	0.564	0.07514	0.588	1.221	2.752	3001	2000

مدل گاما آمیخته



برآورد پارامترها

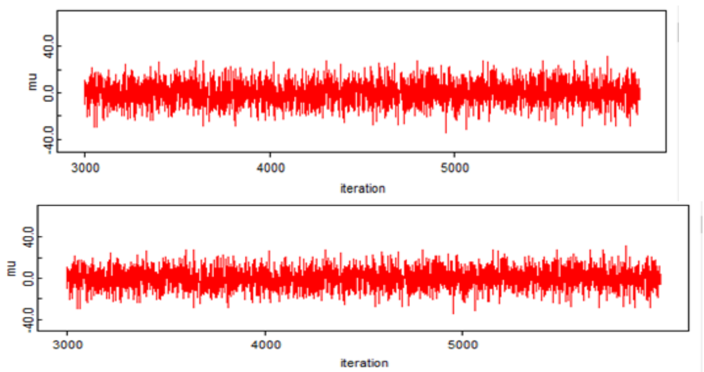
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
a	3.879	19.64	0.4171	0.1066	1.542	18.56	4001	3800

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
b[1]	0.575	4.119	0.08167	0.02423	0.1448	2.934	4001	3800

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
b[1]	0.575	4.119	0.08167	0.02423	0.1448	2.934	4001	3800

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta[1]	5.081	1.191	0.1398	3.097	5.044	7.557	4001	3800

مدل لگ نرمال

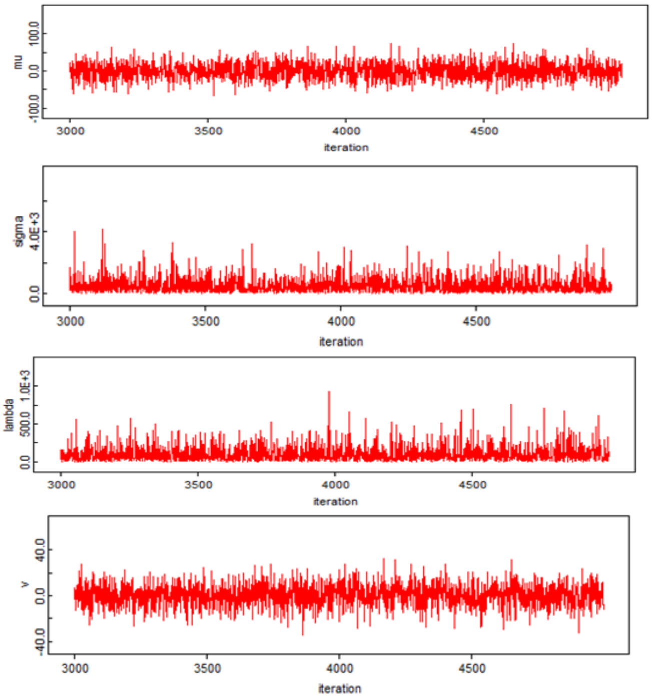


برآورد پارامتر مدل

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu	0.1012	9.943	0.1747	-19.68	0.3077	19.2	3001	3000

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
sigma	100.2	98.04	1.587	2.768	69.47	371.1	3001	3000

مدل وایبل



برآورد پارامترهای مدل

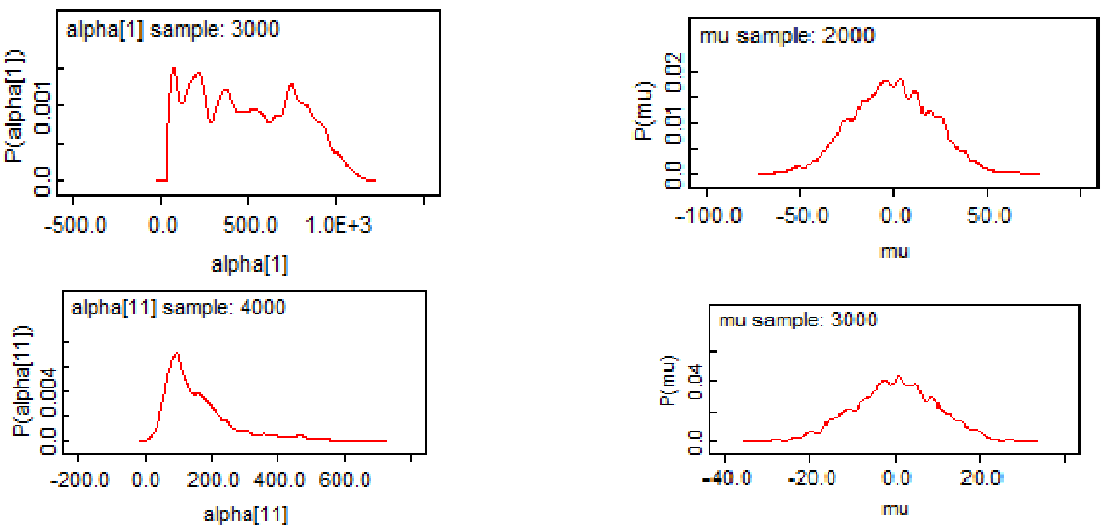
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu	-0.3881	22.22	0.4715	-43.19	-0.7284	43.29	3001	2000

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
sigma	502.7	511.7	11.9	13.37	342.4	1851.0	3001	2000

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
lambda	102.3	103.8	2.026	2.258	72.62	389.8	3001	2000

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
v	0.438	9.737	0.2731	-19.32	0.5791	18.99	3001	2000

چگالی پیشین در چهار مدل



منابع

1. Bali, T. G. (2003). The generalized extreme value distribution. Economics letters, 79(3), 423-427.
2. Balkema, A. A., & De Haan, L. (1974). Residual life time at great age. The Annals of probability, 792-804.
3. Behrens, C. N., Lopes, H. F., & Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. Statistical Modelling, 4(3), 227-244.
4. Carreau, J., & Bengio, Y. (2009). A hybrid pareto mixture for conditional asymmetric fat-tailed distributions. IEEE transactions on neural networks, 20(7), 1087-1101.
5. Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). An introduction to statistical modeling of extreme values (Vol. 208). London: Springer.
6. Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. Journal of the Royal Statistical Society. Series B (Methodological), 363-375.
7. Leadbetter, M., Lindgren, G., & Rootzen, H. (1983). Extreme and related properties of random sequences and series.
8. Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold es-timation and uncertainty quantification. REVSTAT–Statistical Journal, 10(1), 33-60.
9. Titterington, D. M., Smith, A. F., & Makov, U. E. (1985). Statistical analysis of finite mixture distributions. Wiley.