مجله علوم آماری، بهار و تابستان ۱۳۸۸
جلد ۳، شماره ۱، ص ۹۵-۱۰۹

تحلیل بیزی مقادیر کرانگین با استفاده از اسپلاین در مدل آمیختهٔ تعمیم‌یافته

بهزاد محمودیان، موسی گل‌علی‌زاده
گروه آمار، دانشگاه تربیت مدرس

تاریخ دریافت: ۱۳۸۸/۱/۱۵ تاریخ آخرین بازنگری: ۱۳۸۸/۶/۲۷

چکیده : مدل‌بندی پاسخ‌های کرانگین در حضور اثرات غیرخطی، زمانی، فضایی و متقابل می‌تواند با مدل آمیخته صورت پذیرد. به علاوه اسپلاین هموارى در مدل آمیخته و رهیافت بیزی تواماً چارچوب مناسبی را برای استنباط مقادیر کرانگین فراهم می‌کنند. در این مقاله به کارگیری اسپلاین هموارى برای اثر غیرخطی متغیر تبیینی در قالب یک مدل آمیختهٔ تعمیم‌یافته بیان و برای تحلیل مقادیر کرانگین به کار می‌رود. برای این منظور فرض می‌شود که پاسخ‌های کرانگینی مشروط بر تحقق‌هایی از متغیر تبیینی و اثر تصادفی مستقل و دارای توزیع مقدار کرانگین تعمیم‌یافته باشند، سپس پارامتر مکان توزیع به صورت تابع‌ای هموار از متغیر تبیینی با استفاده از تکنیک‌های مونت کارلوی زنجیر مارکوفی در رهیافت بیزی برآورد می‌شود. در پایان مدل ارائه شده برای مدل‌بندی کرانگین داده‌های ازن به کار می‌رود.

آدرس الکترونیک مسئول مقاله: بهزاد محمودیان، mahmoudian@modares.ac.ir
کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۰G۳۲، ۶۲G۰۸

واژه‌های کلیدی : مقادیر کرانگین، توزیع مقدار کرانگین تعمیم‌یافته، اسپلاین هموراری، رهیافت بیزی، مینیماهای بلوکی، داده‌های ازن.

۱ مقدمه

نظریهٔ مقادیر کرانگین [۱] که رفتار مشاهدات خیلی بزرگ یا کوچک را در فرآیندهای تصادفی تحلیل می‌کند، در علوم هیدرولوژی، هواشناسی، مهندسی، زمین‌شناسی و مالی به کار می‌رود. این نظریه توزیع‌ای پارامتری و تقریبی را برای مقادیر کرانگین معرفی می‌کند. در ابتدا به دلیل محدودیت محاسبهٔ برآوردگرها پارامترهای توزیع ثابت فرض می‌شد. بتدریج پارامترها با توابع خطی از متغیرهای تبیینی مدل‌بندی شدند. به عنوان مثال اسمیت (۱۹۸۶) پارامتر مکانی توزیع مقدار کرانگین تعمیم‌یافته را با مولفه‌های خطی و سینوسی مرتبط نمود.

محدودیت‌هایی چون انتخاب فرم تابعی مناسب و استفاده از ضرایب رگرسیونی زیادتر به منظور تضمین انعطاف‌پذیری بیشتر در مدل‌های پارامتری موجب توجه روزافزون به مدل‌های ناپارامتری یا نیمه‌پارامتری با تکیه بر انعطاف‌پذیری در برازش مدل به مشاهدات گردیده است. در این مدل‌ها نوع رابطهٔ بین متغیرهای تبیینی و پاسخ مستقیماً توسط داده‌ها مشخص می‌شود و پارامترهای توزیع نیز با توابعی هموار از متغیرهای تبیینی مدل‌بندی می‌گردند. پاولی و کلز (۲۰۰۱) اسپلاین هموراری [۲]، یی و استفنسون (۲۰۰۷) مدل‌های جمعی تعمیم‌یافتهٔ برداری و پدون و واند (۲۰۰۸) با اسپلاین‌های تاوانیده [۳] در قالب مدل‌های آمیختهٔ تعمیم‌یافته به هموارسازی پاسخ‌های کرانگینی پرداختند.

اجرای مدل‌های فوق برای مقادیر کرانگین به دلیل پیچیده بودن تابع درستنمایی موجب بروز مشکلات محاسباتی جدیدی شد. اسمیت (۱۹۸۵) نشان داد برآوردگرهای ماکسیمم درستنمایی به ازای بعضی از مقادیر پارامتر شکل موجود نیستند. داویسون و رامش (۲۰۰۰) محدودیت برازش مدل با تابع درستنمایی

---

۱ Extreme Value Theory

۲ Smoothing Spline

۳ Penalized Splines

موضعی را برای درجه‌های هموارسازی متفاوت و یی و استفنسون (۲۰۰۷) دشواری محاسبهٔ ماتریس اطلاع فیشر مورد انتظار را از عمدهٔ اشکالات ذکر نموده‌اند. لذا استنباط با رهیافت بیزی به چند دلیل عمده می‌تواند مورد توجه قرار گیرد. در استنباط بیزی تابع درستنمایی شرطی به کار می‌رود و به تابع درستنمایی حاشیه‌ای که معمولاً تقریب زده می‌شود، نیاز نیست. تقریب این حاشیه‌ای‌ها اثر قابل ملاحظه‌ای در دقت برآورد پارامترها به ویژه مولفه‌های واریانس دارد. لارینی و پالی (۲۰۰۹) با رهیافت بیزی کرانگین‌های حاصل از فرآیند نقطه‌ای پواسون را با اسپلاین هموار مدل‌بندی کردند.

در این مقاله اسپلاین هموار برای پارامتر مکان توزیع مینیماهای بلوکی به صورت یک مدل آمیخته مورد استفاده قرار می‌گیرد. در مدل ارائه شده ضرایب اثر تصادفی ثابت نیستند و در طول زمان تغییر می‌کنند. ثابت نبودن ضرایب اثر تصادفی موجب برآورد آن‌ها با توجه به یکدیگر و در نتیجه کاهش بعد اثر تصادفی می‌شود. همچنین یک مولفهٔ خطا برای مدل‌بندی تغییرات داده‌ها که توسط اسپلاین هموار بیان نشده‌اند، در نظر گرفته می‌شود. در ادامه، روش ماکسیمای بلوکی در نظریهٔ مقادیر کرانگین و مدل‌های آمیخته به اختصار در بخش ۲ بیان می‌شوند. سپس در بخش ۳ تحلیل بیزی مقادیر کرانگین با مدل‌های آمیخته ارائه می‌گردد. نحوه به کارگیری مدل ارائه شده در تحلیل مقادیر کرانگین داده‌های ازن شهر تهران در بخش ۴ نشان داده می‌شود. نهایتاً بحث و نتیجه‌گیری در بخش ۵ ارائه می‌گردد.

۲ تحلیل مقادیر کرانگین

در نظریهٔ مقادیر کرانگین معمولاً برای برون‌یابی سطوح کرانگین یک توزیع پارامتری به مقادیر کرانگین برازش داده می‌شود. این توزیع عمدتاً برای محاسبهٔ احتمال تخطی متغیر تصادفی کرانگینی از سرحدی مشخص یا مقداری که این متغیر با احتمال مشخص از آن فزونی می‌یابد استفاده می‌شود. در مدل‌بندی ماکسیمای بلوکی، ابتدا بلوک‌های زمانی به صورت بازه‌های زمانی تعیین و به ماکسیمای بلوک‌ها توزیع پارامتری تقریبی برازش می‌شود. معمولاً انتخاب بلوک‌های زمانی

بزرگ‌تر موجب برازش مناسب‌تر توزیع تقریبی و استقلال ماکسیماها می‌شوند. فرض کنید $\{Z_t\}_{t\geq 1}$ دنباله‌ای از متغیرهای تصادفی پیوسته مستقل و هم‌توزیع و $M_n = \max\{Z_t : t = 1, \ldots, n\}$ باشند. اگر توزیع Z_tها معلوم باشد، توزیع دقیق M_n به راحتی مشخص می‌شود. در غیر این صورت نظریهٔ مقادیر کرانگین برای M_n توزیع تقریبی $P(\frac{M_n - b_n}{a_n} \leq y) \to G(y)$ را به ازای ثابت‌های حقیقی $a_n > 0$ و b_n، زمانی که $n \to \infty$ ارائه می‌کند. اگر $G(\cdot)$ تابع توزیع غیر تباهیده باشد، ثابت می‌شود یکی از توزیع‌های گامبل، فرهشه یا وایبل می‌باشد (کلز، ۲۰۰۱). با پارامترسازی مجدد، این توزیع‌ها را می‌توان به شکلی یکپارچه به نام توزیع مقدار کرانگین تعمیم‌یافته[۴] (GEV) با تابع توزیع

$$G(y; \mu, \sigma, \xi) = \exp\left\{ -\left[1 + \xi(\frac{y-\mu}{\sigma})\right]_+^{-1/\xi} \right\} \tag{1}$$

بر مجموعهٔ $\{y : 1 + \xi/\sigma(y-\mu) > 0\}$ تبدیل نمود، که در آن $(u)_+ = \max\{0, u\}$ و $\mu \in R$، $\sigma > 0$ و $\xi \in R$ به ترتیب پارامترهای مکان، مقیاس و شکل توزیع می‌باشند. در این جا توزیع مقدار کرانگین تعمیم‌یافته با نماد $GEV(\mu, \sigma, \xi)$ نمایش داده می‌شود. تابع توزیع GEV برای مینیماهای بلوکی با تعریف $m_n = \min_{t=1,\ldots,n} Y_t$ و رابطهٔ آن با تابع توزیع ماکسیماها $P(\frac{m_n - d_n}{c_n} \leq y) \to 1 - G(-y)$ بر مجموعهٔ $\{y : 1 - \xi(y-\mu)/\sigma > 0\}$ قابل تعریف است،

$$G(y) = 1 - \exp\left\{ -\left[1 - \xi(\frac{y-\mu}{\sigma})\right]_+^{-1/\xi} \right\} \tag{2}$$

مقدار بازگشت $\frac{1}{p}$ام دوره یا چندک $1 - p$ که هدف برون‌یابی را محقق می‌سازد به دلیل فرم بستهٔ تابع توزیع (۲) از رابطهٔ

$$y_p = \mu + \frac{\sigma}{\xi}[1 - (-\log p)^{-\xi}], \quad \xi \neq 0$$

به دست می‌آید. در ادامه اسپلاین همواری در مدل آمیختهٔ خطی را بررسی نموده و نحوهٔ به کارگیری آن در قالب مدل آمیختهٔ تعمیم‌یافته برای توزیع GEV بیان می‌شود.

---
۴ Generalized Extreme Value Distribution

۱.۲ مدل آمیختهٔ خطی و اسپلاین هموار

مدل رگرسیونی

$$y_i = s(x_i) + \epsilon_i \quad i = 1, \dots, n$$

با خطاهای تصادفی مستقل، هم‌توزیع با $N(0, \sigma_\epsilon^2)$ را برای مشاهدات متعلق به $D \subseteq \Re$ در نظر بگیرید. با فرض آن که $s(\cdot)$ تابع‌ای هموار و انتگرال‌پذیر از توان دوم مشتقات باشد، اسپلاین هموار از کمینه کردن مجموع توان‌های دوم خطاهای تاوانیده به دست می‌آید. اسپلاین هموار رتبه کوچک با توابع پایهٔ شعاعی $\{1, x, |x-a_1|^3, \dots, |x-a_K|^3\}$ و K گره ثابت $a_1 < a_2 < \dots < a_K$ می‌تواند به صورت

$$s(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k |x - a_k|^3 \tag{۳}$$

برای تقریب تابع هموار $s(\cdot)$ به کار رود. اسپلاین هموار با توابع پایهٔ شعاعی به دلیل عدم نیاز به تعداد زیاد گره‌ها برای تقریب تابع هموار در رهیافت بیزی مطلوب می‌باشد (کرینیسیانو و همکاران، ۲۰۰۵). در نتیجه در این مقاله از این توابع پایه استفاده می‌شود. ضرایب رگرسیونی $\boldsymbol{\theta} = (\beta_0, \beta_1, u_1, \dots, u_K)$ در (۳) از کمینه کردن مجموع توان‌های دوم خطای تاوانیده

$$\sum_{i=1}^{n} \{y_i - s(x)\}^2 + \lambda \boldsymbol{\theta}' \boldsymbol{D} \boldsymbol{\theta} \tag{۴}$$

به دست می‌آید، که در آن $\lambda > 0$ پارامتر همواری و

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{o}_{2\times 2} & \boldsymbol{o}_{2\times K} \\ \boldsymbol{o}_{K\times 2} & \boldsymbol{\Omega} \end{pmatrix}$$

ماتریس جریمه، به طوری که در آن $\boldsymbol{o}_{2\times 2}$، $\boldsymbol{o}_{2\times K}$ و $\boldsymbol{o}_{K\times 2}$ ماتریس‌هایی با درآیه‌های صفر و $\boldsymbol{\Omega}$ ماتریس $K \times K$ با درآیهٔ (i,j)ام $|a_i - a_j|$ می‌باشد. اگر $\boldsymbol{y} = (y_1, \dots, y_n)'$ بردار مشاهدات پاسخ و ماتریس‌های طرح $\boldsymbol{X}$ و $\boldsymbol{Z}_K$ به ترتیب با

iامین ردیف $(1 \quad x_i)'$ و $(|x_i-a_1|^3 \dots |x_i-a_K|^3)'$ باشند، رابطهٔ (۴) را می‌توان با تقسیم بر واریانس خطا σ_ϵ^2 به صورت

$$\frac{1}{\sigma_\epsilon^2}||\boldsymbol{y}-\boldsymbol{X}\beta-\boldsymbol{Z}_K\boldsymbol{u}||^2+\frac{\lambda}{\sigma_\epsilon^2}\boldsymbol{u}'\boldsymbol{\Omega}\boldsymbol{u}$$

تعدیل نمود. حال با قرار دادن $\sigma_u^2=\sigma_\epsilon^2/\lambda$، $\boldsymbol{b}=\boldsymbol{\Omega}^{1/2}\boldsymbol{u}$ و $\boldsymbol{Z}=\boldsymbol{Z}_K\boldsymbol{\Omega}^{-1/2}$ و فرض آن که بردار مستقل $(\boldsymbol{u}',\boldsymbol{\epsilon}')'$ دارای توزیع نرمال است، اسپلاین همواری رتبه کوچک در قالب مدل آمیختهٔ خطی به صورت

$$\boldsymbol{y}=\boldsymbol{X}\beta+\boldsymbol{Z}\boldsymbol{b}+\boldsymbol{\epsilon}, \qquad Cov\begin{pmatrix}\boldsymbol{b}\\ \boldsymbol{\epsilon}\end{pmatrix}=\begin{pmatrix}\sigma_b^2\boldsymbol{I}_K & \boldsymbol{o}\\ \boldsymbol{o} & \sigma_\epsilon^2\boldsymbol{I}_n\end{pmatrix}$$

قابل بیان است و توسط بهترین پیش‌بین ناریب خطی یا برآورد شبه درستنمایی تاوانیده به مشاهدات برازش می‌شود. برای مشاهدات غیر گاوسی اسپلاین هموار در صورت امکان از بیشینه کردن تابع درستنمایی تاوانیده به دست می‌آید. با این حال به دلیل مشکل تقریب تابع درستنمایی برای توزیع GEV، رهیافت بیزی با تعیین پیشین‌هایی مناسب برای پارامترهای توزیع و ضرایب رگرسیونی توسط شبیه‌سازی از توزیع پسین به کار می‌رود.

یک روش مناسب برای تعیین گره‌ها استفاده از $a_k=(k+1)/(K+2)$امین چندک نمونه‌ای x_iهای یکتا برای $k=1,\dots,K$ می‌باشد. با تعریف T به عنوان تعداد مقادیر یکتای x_iها، مقدار K از رابطهٔ $K=\max\{5,\min\{\frac{1}{4}\times T,40\}\}$ به دست می‌آید (راپرت، ۲۰۰۲).

۲.۲ مدل آمیختهٔ کرانگینی

مدل آمیختهٔ خطی تعمیم‌یافته برای پارامتر مکان توزیع GEV با اثر تصادفی نرمال به صورت

$$\begin{aligned} Y_i|\boldsymbol{u};\beta,\sigma,\xi &\sim GEV\left(g_i^{-1}\left(\boldsymbol{X}\beta+\boldsymbol{Z}\boldsymbol{u}\right),\sigma,\xi\right), \\ \boldsymbol{u}|\boldsymbol{G} &\sim N(\boldsymbol{o},\boldsymbol{G}), \quad i=1,\dots,n \end{aligned} \tag{5}$$

است، که در آن $\boldsymbol{X}$ و $\boldsymbol{Z}$ ماتریس‌های طرح، β بردار ضرایب اثر خطی، $\boldsymbol{u}$ بردار اثر تصادفی، $\boldsymbol{o}$ بردار با درآیه‌های صفر و $\boldsymbol{G}=\sigma_u^{۲}\boldsymbol{I}$ ماتریس کواریانس می‌باشند. در واقع در مدل آمیختهٔ خطی تعمیم‌یافته فرض می‌شود، Y_i مشروط بر اثر تصادفی $\boldsymbol{u}$ دارای توزیع شرطی $GEV(\mu_i,\sigma,\xi)$ است. در توزیع شرطی پیش‌بین خطی $(\boldsymbol{X\beta}+\boldsymbol{Zu})$ با تابع پیوند $g_i(\boldsymbol{\mu})=(\boldsymbol{X\beta}+\boldsymbol{Zu})_i$ به پارامتر مکان μ_i مرتبط می‌شود. پارامترهای توزیع شرطی GEV در میانگین و واریانس توزیع به صورت

$$\begin{aligned} E(Y_i|\boldsymbol{u}) &= \mu_i+\sigma\big(\Gamma(۱-\xi)-۱\big)/\xi, \\ V(Y_i|\boldsymbol{u}) &= \sigma\big(\Gamma(۱-۲\xi)-\Gamma(۱-\xi)^{۲}\big)/\xi, \quad \xi\neq ۰ \end{aligned}$$

مشارکت می‌کنند. مدل‌بندی مقادیر کرانگین با اسپلاین هموار را می‌توان در قالب مدل آمیختهٔ (۵) به صورت

$$\begin{aligned} (\boldsymbol{X\beta}+\boldsymbol{Zu})_i &= \beta_۰+\beta_۱ x_i+\sum_{k=۱}^{K} u_k|x_i-a_k|^{۳}, \\ \boldsymbol{X} &= \begin{pmatrix} ۱ & x_i \end{pmatrix}, \quad \boldsymbol{Z}_K=\Big(|x_i-a_k|^{۳}\Big)_{۱\le k\le K}, \\ \boldsymbol{\Omega} &= \Big(|a_l-a_k|^{۳}\Big)_{۱\le k,l\le K}, \quad \boldsymbol{Z}=\boldsymbol{Z}_K\boldsymbol{\Omega}^{-۱/۲}, \quad i=۱,\ldots,n \end{aligned}$$

بیان نمود. به منظور برآورد موثری از ضرایب اثر تصادفی این ضرایب را با در نظر گرفتن قدم تصادفی مرتبهٔ اول به صورت

$$u_k=u_{k-۱}+\omega_k; \quad \omega_k\sim N(۰,\sigma_u^{۲}) \qquad (۶)$$

به یکدیگر مرتبط و هموار می‌کنیم. این امر موجب می‌شود در مواردی که روابط پیچیده بین متغیر پاسخ و تبیینی وجود دارد بتوان از تعداد کمتری از گره‌ها استفاده نمود. انتخاب گره‌های بیشتر سبب افزایش بعد اثر تصادفی و وقت‌گیر شدن اجرای الگوریتم مونت کارلوی زنجیر مارکوفی در رهیافت بیزی می‌شود. قدم تصادفی مرتبهٔ اول در رابطهٔ (۶) تغییرات بزرگ $u_k-u_{k-۱}$ را جریمه می‌سازد. قدم تصادفی مرتبهٔ دوم، مدل روند خطی موضعی و مدل اتورگرسیو مرتبهٔ بالاتری را می‌توان برای این ضرایب در نظر گرفت. همچنین برای مدل‌بندی تغییراتی که توسط

اسپلاین همواری بیان نشده‌اند، مولفهٔ خطای مستقل و هم‌توزیع به صورت

$$\mu_i = (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u})_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\mu^2), \quad i = 1, \ldots, n \qquad (7)$$

به مدل اضافه می‌شود. در حالت کلی‌تر می‌توان فرض کرد که مولفهٔ خطا همبستهٔ فضایی یا زمانی است.

۳ تحلیل بیزی مقادیر کرانگین

اگر $\boldsymbol{\nu}' = (\boldsymbol{u}' \;\; \boldsymbol{\epsilon}')$ بردار اثر تصادفی با ماتریس طرح $\boldsymbol{C} = (\boldsymbol{Z} \;\; \boldsymbol{I})$ و تابع چگالی ضرایب اثر تصادفی

$$f(u_k; \sigma_u^2) = (2\pi\sigma_u^2)^{-1/2} \exp\Big\{ -\frac{(u_k - u_{k-1})^2}{2\sigma_u^2} \Big\}, \quad k = 2, \ldots, K$$

و تابع چگالی توزیع شرطی GEV برای مینیماها با مشتق گرفتن از تابع توزیع (۲) به صورت

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\nu}; \sigma, \xi) &= \prod_{i=1}^{n} \frac{1}{\sigma} \Big[1 - \xi \Big(\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{C}\boldsymbol{\nu})_i}{\sigma} \Big) \Big]_+^{-1/\xi - 1} \\ &\quad \times \exp\Big\{ -\Big[1 - \xi \Big(\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{C}\boldsymbol{\nu})_i}{\sigma} \Big) \Big]_+^{-1/\xi} \Big\}, \end{aligned}$$

باشند، برازش مدل و تحلیل مقادیر کرانگین معمولاً با برآورد پارامترها از ماکسیمم کردن تابع درستنمایی

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma, \xi, \sigma_\mu^2; \boldsymbol{y}) &= \int_R f(\boldsymbol{y}|\boldsymbol{\nu}; \sigma, \xi) f(\boldsymbol{\epsilon}|\boldsymbol{u}; \boldsymbol{\beta}, \sigma_\mu^2) f(u_1; \sigma_u^2) \\ &\quad \times \prod_{k=2}^{K} f(u_k; u_{k-1}, \sigma_u^2) d\boldsymbol{\nu}, \end{aligned}$$

انجام می‌گیرد. برآورد درستنمایی پارامترها و $\boldsymbol{\nu}$ نیازمند محاسبه انتگرال‌های پیچیده است. می‌توان با استفاده از رهیافت بیزی و شبیه‌سازی با روش‌های مونت کارلوی زنجیر مارکوفی از توزیع پسین

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma, \sigma_u^2, \sigma_\mu^2, \xi|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\nu}; \sigma, \xi) f(\boldsymbol{\epsilon}|\boldsymbol{u}; \boldsymbol{\beta}, \sigma_\mu^2) \prod_{k=2}^{K} f(u_k; u_{k-1}, \sigma_u^2) \\ &\quad \times f(u_1; \sigma_u^2) \pi(\boldsymbol{\beta}) \pi(\sigma) \pi(\xi) \pi(\sigma_\mu^2) \pi(\sigma_u^2) \qquad (8) \end{aligned}$$

بر این مشکل فائق آمد. با تعیین پیشین‌هایی مناسب برای پارامترها، مدل اسپلاین کرانگینی به وسیلهٔ الگوریتم متروپولیس-هستینگس به مشاهدات برازش می‌شود. با تصریح تابع چگالی شرطی مشاهدات و اثر تصادفی برای انجام استنباط بیزی لازم است مسئلهٔ انتخاب پیشین برای پارامترها مورد بررسی قرار گیرد.

وقتی اطلاعات پیشینی در مورد پارامترها در اختیار نباشد و استنباط با رهیافت بیزی صرفاً برای برطرف نمودن محدودیت‌های محاسباتی استفاده شود معمولاً غلبهٔ اطلاعات مشاهدات در تابع درستنمایی بر اطلاعات پیشین ارجحیت دارد. در نتیجه از پیشین‌های ناآگاهی‌بخش استفاده می‌شود. توزیع پیشین نرمال $N(\boldsymbol{o}, \boldsymbol{F})$ برای ضرایب اثر ثابت به ازای ماتریس کواریانس $\boldsymbol{F}$ انتخاب معمول می‌باشد، که در آن معمولاً $\boldsymbol{F}$ ماتریس قطری با واریانس‌های بزرگ اختیار می‌شود. در واقع توزیع پیشین مبهم برای ضرایب اثر ثابت انتخاب می‌شود.

همواری یک هموارساز در رهیافت بسامدی به انتخاب مقدار پارامتر همواری بستگی دارد، در رهیافت بیزی این همواری توسط واریانس $\sigma_u^۲$ کنترل می‌شود. پس انتخاب پیشین برای $\sigma_u^۲$ بر همواری مدل اسپلاینی تاثیرگذار است. توزیع گامای معکوس برای پارامتر واریانس اثر تصادفی، $IG(a,b)$ با انتخاب‌های استاندارد $a = b = \epsilon$ یا $a = ۱, b = \epsilon$ می‌تواند به کار رود که در آن $\epsilon = ۱, ۰/۱, ۰/۰۱, ۰/۰۰۱$ باشد. مدل اسپلاینی حساسیت چندانی به انتخاب توزیع پیشین برای پارامترهای مقیاس و شکل توزیع GEV ندارد و می‌توان توزیع مبهمی را برای آن‌ها در نظر گرفت.

برای شبیه‌سازی از توزیع پسین (۸) در ابتدا پارامتر شکل و مقیاس به ترتیب و جداگانه با گام‌های متروپولیس-هستینگس بهنگام می‌شوند. توزیع‌های شرطی کامل پارامترهای مقیاس و شکل با انتخاب پیشین‌های $N(m_\sigma, s_\sigma^۲)$ و $N(m_\xi, s_\xi^۲)$ با حاصل‌ضرب توزیع پیشین در تابع درستنمایی به دست می‌آیند. به همین ترتیب $\boldsymbol{\beta} = (\beta_۰, \beta_۱)$، $\boldsymbol{u}$ و $\boldsymbol{\epsilon}$ که دارای پیشین نرمال‌اند جداگانه با گام‌های الگوریتم متروپولیس-هستینگس بهنگام می‌شوند. انتخاب توزیع پیشین مزدوج گامای معکوس $IG(\alpha_u, \gamma_u)$ و $IG(\alpha_\mu, \gamma_\mu)$ به ترتیب برای $\sigma_u^۲$ و $\sigma_\mu^۲$ موجب سادگی توزیع شرطی کامل این پارامترها می‌شود.

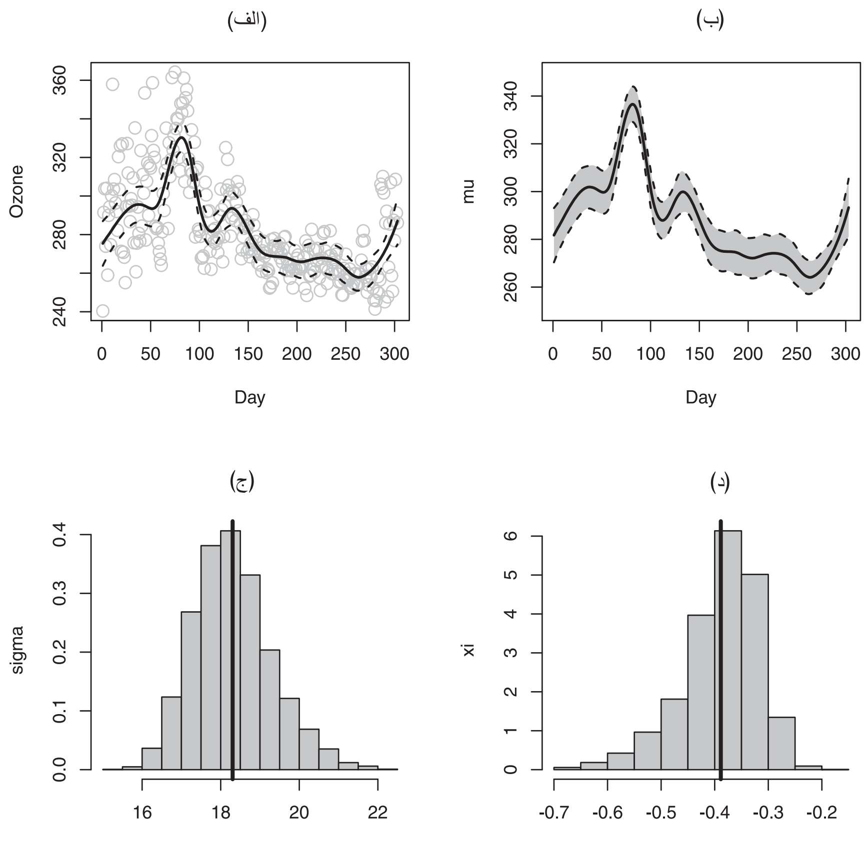
نمونه‌گیری از توزیع شرطی کامل σ_u^2 و σ_μ^2 به ترتیب با شبیه‌سازی از توزیع گامای $\tau_u \sim G(\alpha_u + \frac{K}{2}, \gamma_u + \frac{\|\boldsymbol{u}\|^2}{2})$ و $\tau_\mu \sim G(\alpha_\mu + \frac{n}{2}, \gamma_\mu + \frac{\|\boldsymbol{\mu}-\boldsymbol{X\beta}-\boldsymbol{Zu}\|^2}{2})$ و بهنگام‌سازی آن‌ها با در نظر گرفتن تبدیل $\sigma_u^2 = 1/\tau_u$ و $\sigma_\mu^2 = 1/\tau_\mu$ انجام می‌گیرد. پارامترهای توزیع گاما به گونه‌ای است که $G(a,b)$ دارای میانگین a/b و واریانس a/b^2 است. توزیع پیشین گامای معکوس به شدت به مقیاس مشاهدات حاصل از متغیرهای تبیینی و پاسخ حساسیت دارد (کریسینیانو و همکاران، ۲۰۰۵). این حساسیت در تحلیل مقادیر کرانگین به دلیل بزرگ بودن پراکندگی داده‌ها بیشتر نیز می‌شود. در صورت استاندارد کردن متغیرهای پاسخ و تبیینی تا اندازه‌ای این حساسیت کاهش می‌یابد. استاندارد کردن متغیرها هنگام مدل‌بندی داده‌های واقعی در مقایسه با داده‌های شبیه‌سازی شده به دلیل متفاوت بودن مقیاس هر یک از متغیرها از اهمیت بیشتری برخوردار است (گلمن، ۲۰۰۶). شبیه‌سازی از توزیع پسین پارامترها با گام‌های الگوریتم متروپولیس-هستینگس مشروط بر مقادیر اولیه و ابرپارامترها در تکرار $t+1$ شامل شبیه‌سازی از توزیع‌های شرطی کامل پارامترهای ξ، σ، $\boldsymbol{\nu}$، ϵ و توزیع گامای معکوس برای مولفه‌های واریانس می‌باشد.

پارامتر همواری را می‌توان به صورت $\lambda = \frac{\sigma_u^2}{\sigma_\mu^2}$ با استفاده از نمونه‌های شبیه‌سازی شده به دست آورد. مقادیر کوچک (بزرگ) λ بیش همواری (کم همواری) مدل برازش شده را بیان می‌کند.

۴ تحلیل مقادیر کرانگین ازن شهر تهران

در این بخش اثر متغیر تبیینی زمان بر مینیماهای روزانهٔ ازن شهر تهران با در نظرگرفتن مدل (۷) بررسی می‌شود. ازن (O_3) در اثر فرآیندهای فتوشیمیایی متفاوت در جو حاصل می‌شود. این فرآیندها به شدت تحت تاثیر شرایط اقلیمی حاکم بر منطقهٔ جغرافیایی هستند. به عنوان مثال انتظار می‌رود ازن در یک روز آفتابی بدون باد در تابستان دارای غلظت بیشتری باشد (اسمیت، ۲۰۰۴). ازن به صورت ناهمگن در لایه‌های استراتوسفر و تروپوسفر جو پراکنده شده است. از آن جایی که قسمت اعظم ازن جو در استراتوسفر واقع شده است این ناحیه از ازن، لایهٔ

ازن استراتوسفری نامیده می‌شود. ازن تروپوسفری و ازن استراتوسفری دو نقش متفاوت را در جو از خود بروز می‌دهند. ازن استراتوسفری که مانع تابش اشعه‌های مضر خورشید شده ازن مفید نامیده می‌شود. در مقابل ازن مفید، ازن تروپوسفری، گاز سمی است که در طول روز در نواحی آلوده مثل مناطق شهری تولید می‌شود. داده‌های ازن با واحد دابسن طی چندین نوبت در روز اندازه‌گیری و ثبت می‌شوند[۵].



شکل ۱: (الف) نمودار پراکنش و منحنی برآورد شده، (ب) مدل اسپلاین برآورد شده (خط‌پر) با فاصلهٔ اطمینان ۹۵٪ و (ج) و (د) بافت‌نگار توزیع پسین پارامترهای مقیاس و شکل.

۵ http://geophysics.ut.ac.ir/En/Section/Ozone/Dobson

داده‌های ازن شهر تهران در سال ۲۰۰۹ برای استخراج مینیماهای روزانه مورد استفاده قرار می‌گیرد. به دلیل آن که مقادیر ازن کمتر از مقدار مرزی ۲۲۰ واحد دابسن می‌تواند نشانهٔ وجود آلودگی و تهدیدی برای نازک شدن لایه ازن باشد، برای تحلیل آماری پاسخ‌های کرانگینی، مینیمای روزانهٔ این داده‌ها در نظر گرفته می‌شود. همچنین وجود داده‌های گمشده موجب شد که در مجموع ۳۰۳ مینیمای روزانهٔ استخراج شود. بنابراین برای بررسی تغییرات ازن مفید جو و تعیین روند آن در حضور متغیر تبیینی زمان مینیماهای روزانهٔ ازن شهر تهران با استفاده از مدل آمیختهٔ خطی که در رابطهٔ (۷) بیان شده است با $K = ۲۰$ گره مدل‌بندی می‌شوند. برای برازش مدل به مشاهدات، توزیع‌های پیشین مبهم

$$\beta_0,\ \beta_1 \sim N(0, 10^3),\ \ \log(\sigma),\ \xi \sim N(0, 10^3),\ \ \sigma_u^2,\ \sigma_\mu^2 \sim IG(10^{-6}, 10^{-6})$$

انتخاب شده‌اند. پارامتر میزان‌سازی در الگوریتم متروپولیس-هستینگس به گونه‌ای تنظیم گردید که نرخ پذیرش زنجیر شبیه‌سازی شدهٔ پارامترها در حدود ۳۰ تا ۴۰ درصد باشند. استنباط بیزی براساس نمونه‌ای به حجم ۵۰۰۰ از زنجیر شبیه‌سازی شده با تکرار $۱/۵ \times ۱۰^۵$، مرحلهٔ داغیدن ۵×۱۰^۴ و تاخیر بیستم می‌باشد. شکل ۱ نمودارهای پراکنش، برآورد مدل اسپلاین در پارامتر مکان با فاصله اطمینان ۹۵٪ و بافت‌نگار پارامترهای مقیاس و شکل توزیع GEV را نشان می‌دهد. منحنی مدل برآورد شده به وسیلهٔ میانگین توزیع پسین و نوار اطمینان باریک بیزی در نمودار پراکنش به ترتیب بیانگر برازش مناسب به مشاهدات و دقت بالای برآورد پارامترها است. در نمودار تابع اسپلاین برآوردی پارامتر مکان، تغییرات مشابه با نمودار پراکنش را می‌توان مشاهده نمود. بافت‌نگار توزیع پسین پارامترهای ξ و σ در شکل ۱ (ج و د) برآورد بیزی با میانگین توزیع پسین (خط عمودی)، توزیع متقارن و دقت مناسب این پارامترها را نشان می‌دهد. مقدار منفی برآورد پارامتر ξ نشان می‌دهد که توزیع GEV توزیع وایبل معکوس و از پایین کراندار است. همان طور که ملاحظه می‌شود مینیماهای روزانهٔ ازن با روند کاهشی همراه است که در روزهای پایانی سال به کمترین مقدار خود می‌رسد. با توجه به مدل برازش شده و در اختیار داشتن

توزیع پسین پارامترها می‌توان پیشگویی مینیمای ازن و احتمال پیشامدهای دلخواه را در زمان مشخص محاسبه نمود. به عنوان مثال پیشگویی مقدار مینیمای روزانهٔ ازن در روز صد و پنجاهم سال **۲۷۹/۴۱** واحد دابسن با انحراف استاندارد **۱۰/۵۲** به دست آمد. خطای برآورد با توجه به پراکندگی مینیماهای ازن خطای کوچکی محسوب می‌شود. همچنین احتمال این که در روز صد و پنجاهم سال مقدار مینیمای روزانهٔ ازن از ۲۵۰ واحد دابسن کمتر باشد، ۰/۰۰۱ برآورد شد.

۵ بحث و نتیجه‌گیری

در این مقاله مدل آمیختهٔ خطی تعمیم‌یافته با استفاده از اسپلاین هموار برای تحلیل داده‌های ازن با رهیافت بیزی به کار گرفته شد. بیان اسپلاین هموار با مدل آمیخته‌ای که ضرایب اثر تصادفی آن ثابت نیستند و انجام استنباط با رهیافت بیزی برای مدل‌بندی وابستگی بین متغیرهای پاسخ و تبیینی به راحتی انجام گرفت. در حالی که در رهیافت بسامدی نیاز به روشی برای برآورد پارامتر هموار و تعیین درستی آن است، این پارامتر به صورت تقسیم دو مولفهٔ واریانس در رهیافت بیزی محاسبه گردید. همچنین موجود نبودن برآوردگرهای ماکسیمم درستنمایی به ازای بعضی از مقادیر پارامتر شکل انگیزه‌ای برای استفاده از رهیافت بیزی در این مقاله را فراهم نمود. می‌توان تعمیم‌هایی را برای مدل در نظر گرفته شده بیان نمود. اولاً فرض توزیع نرمال برای اثر تصادفی و مولفهٔ خطا را می‌توان با توزیع‌هایی نظیر t-استیودنت، نمایی دوپارامتری و یا توزیع چوله یا دم کلفت‌تر جایگزین کرد. هر کدام از توزیع‌های بیان شده چارچوب استواری را برای برآورد پارامترها فراهم می‌کنند. در نظر گرفتن ساختار همسبتگی فضایی یا زمانی برای مولفهٔ خطا و در نظر گرفتن مدل مشابه برای پارامترهای مقیاس و شکل توزیع مقدار کرانگین تعمیم‌یافته از مسائل قابل بررسی دیگر می‌باشند.

تقدیر و تشکر

از پیشنهادات ارزنده و نظرات اصلاحی داوران محترم مجله صمیمانه تشکر می‌شود.

مراجع

Coles, S. G., (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.

Coles, S. G. and Powell, E. A. (1996), Bayesian Methods in Extreme Value Modeling: a Review and new Development, *International Statistical Review*, **64**, 119-136.

Crainiceanu, C. M., Ruppert, D. and Wand, M. P. (2005), Bayesian Analysis for Penalized Spline Regression Using WinBUGS, *Journal of Statistical Software*, **14**, 1-24.

Davison, A. C. and Ramesh, N. I. (2000), Local Likelihood Smoothing of Sample Extremes, *Journal of the Royal Statistical Society, Series B.*, **62**, 191-208.

Gelman, A. (2006), Prior Distributions for Variance Parameters in Hierarchical Model, *Bayesian Analysis*, **1**, 515-533.

Laurini, F. and Pauli, F. (2009), Smoothing Sample Extremes: the Mixed Model Approach, *Computational Statistics and Data Analysis*, **53**, 3842-3854.

Padoan, S. A. and Wand, M. P. (2008), Mixed Model-based Additive Models for Sample Extremes, *Statistics and Probability Letters*, **78**, 2850-2858.

Pauli, F. and Coles, S. G. (2001), Penalized Likelihood Inference in Extreme Value Analysis, *Journal of Applied Statistics*, **28**, 547-560.

Ruppert, D. (2002), Selecting the Number of Knots for Penalized Splines, *Journal of Computational and Graphical Statistics*, **11**, 735-757.

Smith, R. L. (1985), Maximum Likelihood Estimation in a Class of Nonregular Cases, *Biometrika*, **72**, 67-90.

Smith, R. L. (1986), Extreme Value Theory Based on the r-largest Annual Events, *Journal of Hydrology*, **86**, 27-43.

Smith, R. L. (2004), Statistics of Extremes, with Applictions in Environment, Insurance and Finance. In Finkenstadt, B. and Rootzen, H., editor, *Extreme Values in Finance, Telecomunications and the Environment*, 1-78, Chapman and Hall, London.

Yee, T. W. and Stephenson, A. G. (2007), Vector Generalized Linear and Additive Extreme Value Models, *Extremes*, **10**, 1-19.