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

بهزاد محمودیان، محسن محمدزاده، موسی گل‌علی‌زاده
گروه آمار، دانشگاه تربیت مدرس

مدل‌بندی پاسخ‌های کرانگین در حضور اثرات خطی، غیرخطی، زمانی، فضایی و متقابل می‌تواند با مدل آمیخته صورت پذیرد. بعلاوه اسپلاین هموراری در مدل آمیخته و رهیافت بیزی تواماً چارچوب مناسبی را برای استنباط مقادیر کرانگین فراهم می‌کنند. در این مقاله بکارگیری اسپلاین هموراری برای اثرات غیرخطی در قالب یک مدل آمیختهٔ تعمیم‌یافته بیان و برای تحلیل مقادیر کرانگین بکار می‌رود. برای این منظور فرض می‌شود که پاسخ‌های کرانگینی مشروط بر تحققهایی از متغیرهای تبیینی و اثرات تصادفی مستقل و دارای توزیع مقدار کرانگین تعمیم‌یافته باشند، سپس پارامترهای توزیع بصورت توابعی هموار از متغیرهای تبیینی با استفاده از تکنیک‌های زنجیر مارکف مونت‌کارلویی و نمونه‌گیر برشی در رهیافت بیزی برآورد می‌شوند. در پایان مدل ارائه شده برای مدل‌بندی کرانگین داده‌های ازن بکار می‌رود.

واژه‌های کلیدی: مقادیر کرانگین، توزیع مقدار کرانگین، اسپلاین هموراری، زنجیر مارکوف مونت کارلویی، مینیماهای بلوکی، داده‌های ازن.

۱ مقدمه

نظریهٔ مقادیر کرانگین [1] که رفتار مشاهدات خیلی بزرگ یا کوچک را در فرآیندهای تصادفی تحلیل می‌کند، در علوم هیدرولوژی، هواشناسی، مهندسی، زمین‌شناسی و مالی بکار می‌رود. این نظریه توزیع‌ای پارامتری و تقریبی را برای مقادیر کرانگین معرفی می‌کند. در ابتدا بدلیل محدودیت محاسبهٔ برآوردگرها پارامترهای توزیع ثابت فرض می‌شد. بتدریج پارامترها با توابع خطی از متغیرهای تبیینی مدل‌بندی شدند. بعنوان مثال اسمیت (۱۹۸۶) پارامتر مکانی

[1] Extreme value theory

Archive of SID
www.SID.ir

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 هموار، یی و استفنسون (۲۰۰۷) مدل‌های جمعی تعمیم‌یافتهٔ برداری و پدون و واند (۲۰۰۸) با اسپلاین‌های تاوانیده در قالب مدل‌های آمیختهٔ تعمیم‌یافته به هموارسازی پاسخ‌های کرانگینی پرداختند. اجرای مدل‌های فوق برای مقادیر کرانگین بدلیل پچیده 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 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 بکار می‌رود و به تابع درستنمایی حاشیه‌ای که معمولاً تقریب زده می‌شود، نیاز نیست. تقریب این حاشیه‌ای‌ها اثر قابل ملاحظه‌ای در دقت برآورد پارامترها بویژه مولفه‌های واریانس دارد. از طرف دیگر مقادیر کرانگین از تغییرپذیری زیادی برخوردارند، در نتیجه استفاده از اطلاعاتی در قالب توزیع پیشین می‌تواند موجب افزایش دقت نتایج گردد. لارینی و پالی (۲۰۰۹) با رهیافت بیزی کرانگین‌های حاصل از فرآیند نقطه‌ای پواسن را با اسپلاین‌های هموار مدل‌بندی کردند. در این مقاله هموارساز اسپلاین برای پارامتر مکانی توزیع مینیماهای بلوکی بصورت یک مدل آمیخته مورد استفاده قرار می‌گیرد. سپس نتایج استنباط بیزی با حالتی که پارامترهای توزیع ثابت یا تبدیلی خطی از متغیرهای تبیینی است، مقایسه می‌شود. در ادامه، روش ماکزیمای 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 بکارگیری 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.

www.SID.ir

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 ماکزیمای بلوکی، ابتدا بلوک‌های زمانی بصورت بازه‌های زمانی تعیین و به ماکزیمای بلوک‌ها توزیع پارامتری تقریبی برازش می‌شود. معمولاً انتخاب بلوک‌های زمانی بزرگتر موجب برازش مناسب‌تر توزیع تقریبی و استقلال ماکزیماها 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_i ها معلوم باشد، توزیع دقیق M_n براحتی 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، زمانیکه $n \to \infty$ ارائه می‌کند. اگر $G(\cdot)$ تابع توزیع غیر تباهیده باشد، ثابت می‌شود یکی از توزیع‌های گامبل، فرهشه یا وایبل می‌باشد (کلز، ۲۰۰۱). با بازپارامتریدن پارامترها، این توزیع‌ها را می‌توان به شکلی یکپارچه بنام توزیع مقدار کرانگین تعمیم‌یافته [2] (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 $\xi \in R$ بترتیب پارامترهای مکانی، مقیاس و شکل توزیع می‌باشند. در اینجا توزیع مقدار کرانگین تعمیم‌یافته با نماد $GEV(\mu,\sigma,\xi)$ نمایش داده می‌شود. ویژگی مفید و خروجی نظریهٔ مقادیر کرانگین، مقادیر بازگشت y_p با دورهٔ بازگشت $1/p$ است، که از تابع توزیع (۱) بصورت

$$y_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - (-\log(1-p))^{-\xi}], & \xi \neq 0 \\ \mu - \sigma\log(-\log(1-p)), & \xi = 0 \end{cases}$$

بدست می‌آید. همچنین مقادیر بازگشت برای توزیع مینیماهای بلوکی را می‌توان با تعریف $m_n = \min\{Z_t : t = 1, \ldots, n\}$ و رابطهٔ آن با تابع توزیع ماکزیماها

[2] Generalized extreme value distribution

Archive of SID
www.SID.ir

$P(\frac{m_n-d_n}{c_n} \leq y) \to 1 - G(-y)$ بصورت

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

بدست آورد (کلز، ۲۰۰۱). در ادامه اسپلاین هموار در مدل آمیختهٔ خطی را بررسی نموده و نحوهٔ بکارگیری آن در قالب مدل آمیختهٔ تعمیم‌یافته برای توزیع GEV بیان می‌شود.

۲-۱ مدل آمیختهٔ خطی و اسپلاین هموار

مدل رگرسیونی

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

با خطاهای تصادفی مستقل، هم‌توزیع با $N(0, \sigma_\epsilon^2)$ را برای مشاهدات متعلق به $D \subseteq \Re$ در نظر بگیرید. با فرض آنکه $s(\cdot)$ تابع‌ای هموار و انتگرال‌پذیر از توان دوم مشتقات باشد، اسپلاین هموار از کمینه کردن مجموع توان‌های دوم خطاهای تاوانیده بدست می آید. بدلیل خواص آمیختگی مناسب زنجیر مارکف مونت کارلویی در رهیافت بیزی اسپلاین هموار[3] رتبه کوچک برای مدل‌بندی تابع هموار $s(\cdot)$ انتخاب می‌شود (کرینیسیانو و همکاران، ۲۰۰۵). اسپلاین هموار رتبه کوچک با توابع پایه $\{1, x, |x-a_1|^3, \ldots, |x-a_K|^3\}$ و K گره ثابت $a_1 < a_2 < \ldots < a_K$ بصورت

$$s(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k |x - a_k|^3 \qquad (2)$$

است. ضرایب رگرسیونی $\boldsymbol{\theta} = (\beta_0, \beta_1, u_1, \ldots, u_K)$ در (۲) از کمینه کردن مجموع توان‌های دوم خطای تاوانیده

$$\sum_{i=1}^{n} \{y_i - s(x)\}^2 + \lambda \boldsymbol{\theta}' \boldsymbol{D} \boldsymbol{\theta} \qquad (3)$$

بدست می آید، که در آن $\lambda > 0$ پارامتر هموار و

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

[3] Smoothing spline

Archive of SID
www.SID.ir

ماتریس تاوانیده است. بطوریکه $\boldsymbol{o}_{2\times 2}$، $\boldsymbol{o}_{2\times K}$ و $\boldsymbol{o}_{K\times 2}$ ماتریس‌هایی با درآیه‌های صفر و $\boldsymbol{\Omega}$ ماتریس $K \times K$ با درآیهٔ (i,j)ام $|a_i - a_j|$ است. اگر $\boldsymbol{y} = (y_1, \dots, y_n)'$ بردار مشاهدات پاسخ و ماتریس‌های طرح $\boldsymbol{X}$ و $\boldsymbol{Z}_K$ بترتیب با i امین ردیف $(1 \quad x_i)'$ و $\left(|x_i - a_1|^3, \dots, |x_i - a_K|^3\right)'$ باشند، رابطهٔ (۳) را می‌توان با تقسیم بر واریانس خطا σ_ϵ^2 بصورت زیر تعدیل نمود:

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

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 آنکه بردار مستقل $(\boldsymbol{u}', \boldsymbol{\epsilon}')'$ دارای توزیع نرمال است، اسپلاین هموار رتبه کوچک در قالب مدل آمیختهٔ خطی بصورت

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

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 لگاریتم درستنمایی تاوانیدهٔ بدست می‌آید. با این حال بدلیل مشکل تقریب تابع درستنمایی برای مقادیر کرانگین، رهیافت بیزی با تعیین پیشین‌هایی مناسب برای پارامترهای توزیع و ضرایب رگرسیونی توسط شبیه‌سازی از توزیع پسین بکار می‌رود.

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

مدل آمیخته برای پارامتر مکانی توزیع GEV با اثرات تصادفی گاوسی بصورت

$$y_i|\boldsymbol{\beta}, \boldsymbol{u}, \sigma, \xi \sim GEV(\mu_i = (\boldsymbol{X\beta} + \boldsymbol{Zu})_i, \sigma, \xi), \quad \boldsymbol{u}|\boldsymbol{G} \sim N(\boldsymbol{o}, \boldsymbol{G}), \quad i = 1, \dots, n \quad (4)$$

است، که در آن $\boldsymbol{X}$ و $\boldsymbol{Z}$ ماتریس‌های طرح، $\boldsymbol{\beta}$ بردار ضرایب اثرات خطی، $\boldsymbol{u}$ بردار اثرات تصادفی، $\boldsymbol{o}$ بردار با درآیه‌های صفر و $\boldsymbol{G} = \sigma_u^2\boldsymbol{I}$ ماتریس کواریانس می‌باشند. مدل‌بندی مقادیر کرانگین با اسپلاین هموار را می‌توان در قالب مدل آمیختهٔ (۴) بصورت زیر بیان نمود:

$$(\boldsymbol{X\beta} + \boldsymbol{Zu})_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} u_k|x_i - a_k|^3, \qquad i = 1, \dots, n$$

Archive of SID
www.SID.ir

$$\boldsymbol{X} = \begin{pmatrix} ۱ & x_i \end{pmatrix}_{۱\leq i\leq n}, \quad \boldsymbol{Z} = \left[|x_i - a_k|^{۳} \right]_{۱\leq i\leq n, ۱\leq k\leq K},$$
$$Cov(\boldsymbol{u}) = \sigma_u^{۲}\boldsymbol{I}_K$$

۳ تحلیل بیزی مقادیر کرانگین

اگر $\boldsymbol{\nu}' = [\beta' \,,\, \boldsymbol{u}']$ بردار اثرات ثابت و تصادفی با ماتریس طرح $\boldsymbol{C} = [\boldsymbol{X} \,,\, \boldsymbol{Z}]$ و توابع چگالی اثرات تصادفی و مشاهدات بترتیب

$$f(\boldsymbol{u};\sigma_u^{۲}) = (۲\pi\sigma_u^{۲})^{-K/۲}\exp\Big\{-\frac{||\boldsymbol{u}||^{۲}}{۲\sigma_u^{۲}}\Big\},$$

و

$$f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\beta},\sigma,\xi) = \prod_{i=۱}^{n}\frac{۱}{\sigma}\Big\{۱-\xi\Big(\frac{(\boldsymbol{y}-\boldsymbol{C\nu})_i}{\sigma}\Big)\Big\}^{-۱/\xi-۱} \times\exp\Big\{-\Big\{۱-\xi\Big(\frac{(\boldsymbol{y}-\boldsymbol{C\nu})_i}{\sigma}\Big)\Big\}^{-۱/\xi}\Big\},$$

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

$$L(\boldsymbol{\beta},\sigma,\xi,\sigma_u^{۲}) = f(\boldsymbol{y};\boldsymbol{\beta},\sigma,\xi,\sigma_u^{۲}) = \int_R f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\beta},\sigma,\xi)f(\boldsymbol{u};\sigma_u^{۲})d\boldsymbol{u},$$

و پیش‌بینی اثرات تصادفی با پیش‌بین $\hat{\boldsymbol{u}} = E(\boldsymbol{u}|\boldsymbol{y})$ انجام می‌گیرد (پدون و واند، ۲۰۰۸). برآورد درستنمایی پارامترها و u نیازمند محاسبه انتگرال‌های پیچیده است. می‌توان با استفاده از رهیافت بیزی و شبیه‌سازی با زنجیر مارکف مونتِ کارلویی از توزیع پسین

$$p(\boldsymbol{\beta},\boldsymbol{u},\sigma,\sigma_u^{۲},\xi|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\beta},\sigma,\xi)f(\boldsymbol{u};\sigma_u^{۲})\pi(\boldsymbol{\beta})\pi(\sigma)\pi(\xi)\pi(\sigma_u^{۲})$$

بر این مشکل فائق آمد. با تعیین پیشین‌هایی مناسب برای پارامترها، اسپلاین هموار با شبیه‌سازی زنجیر مارکف بوسیلهٔ الگوریتم نمونه‌گیری برشی[۴] به مشاهدات برازش می‌شود. با

[4] Slice sampling

Archive of SID
www.SID.ir

تصریح تابع چگالی مشاهدات و اثرات تصادفی، برای انجام استنباط بیزی لازم است مسئله انتخاب پیشین برای پارامترها مورد بررسی قرار گیرد.

وقتی اطلاعات پیشینی در مورد پارامترها در اختیار نباشد و روش‌های زنجیر مارکف مونت کارلویی صرفاً برای برطرف نمودن محدودیت‌های محاسباتی بکار رود، معمولاً غلبهٔ اطلاعات مشاهدات در تابع درستنمایی بر اطلاعات پیشین‌ها ارجحیت دارد. در نتیجه از پیشین‌های ناآگاهی‌بخش استفاده می‌شود. با این حال زمانیکه تابع درستنمایی اطلاعات دقیقی را فراهم نکند حتی یک پیشین ناآگاهی‌بخش تاثیر بسزایی بر استنباط دارد (لمبرت و همکاران، ۲۰۰۵). توزیع نرمال $N(\boldsymbol{o},\boldsymbol{F})$، که در آن ماتریس کواریانس $\boldsymbol{F}$ قطری با واریانس‌های بزرگ است، یک انتخاب معمول به عنوان توزیع پیشین ناآگاهی‌بخش برای ضرایب اثرات ثابت می‌باشد. توزیع یکنواخت $U(-A,A)$ نیز با مقادیر بزرگ A برای مبهم بودن توزیع پیشین انتخابی دیگر برای β ها است (گلمن، ۲۰۰۶).

انتخاب پیشین برای پارامترهای واریانس از اهمیت بیشتری برخوردار است. لمبرت و همکاران (۲۰۰۵) برای پارامتر مقیاس در مدل سلسله مراتبی عملکرد توزیع‌های یکنواخت، نیمه نرمال، گاما، پارتو و لجستیک را با مطالعهٔ شبیه‌سازی مقایسه نمودند. آنها دریافتند که هیچ‌کدام از این توزیع‌های پیشین دارای نتیجهٔ بهتری در شرایط متفاوت نبوده اما با افزایش واحدها در مدل سلسله مراتبی اثر انتخاب نوع پیشین کاهش می‌یابد. وقتی اثرات تصادفی دارای توزیع نرمال هستند، توزیع گامای معکوس، $IG(a,b)$، با انتخاب‌های استاندارد $a=b=\epsilon$ یا $a=1, b=\epsilon$ برای مقادیر $\epsilon = 1, 0/1, 0/01, 0/001$ انتخاب دیگری است (فهرمیر و لانگ، ۲۰۰۱). برای مقادیر خیلی کوچک ϵ توزیع پسین پارامترها ناسره خواهد شد. گلمن (۲۰۰۶) توزیع t ناقص را ضمن مقایسه با توزیع یکنواخت و گامای معکوس به عنوان توزیع نیمه آگاهی‌بخش پیشنهاد کرد. خانوادهٔ توزیع tی ناقص با پارامتر مقیاس B و درجهٔ آزادی ν بصورت:

$$p(\sigma^2) \propto (1 + \frac{1}{\nu}(\frac{\sigma}{B})^2)^{-(\nu+1)/2}$$

است، که به ازای $\nu = -1$ و $\nu = 1$ بترتیب توزیع‌های یکنواخت و کشی سره را نتیجه می‌دهد.

از آنجایی که توزیع پیشین مزدوج برای پارامترهای توزیع مقدار کرانگین وجود ندارد (کلز و پاول، ۱۹۹۶)، توزیع‌های پیشین یکنواخت، نرمال بریده شده و بتا برای پارامتر شکل و توزیع کشی از خانوادهٔ توزیع‌های tی ناقص (گلمن، ۲۰۰۶) و گامای معکوس برای پارامتر مقیاس انتخاب می‌شود. همچنین می‌توان پارامتر مقیاس را نیز مستقل از پارامتر مکانی بر حسب متغیرهای تبیینی مدل‌بندی نمود.

Archive of SID
www.SID.ir

۴ تحلیل کرانگین‌های ازن در تهران

در این بخش کاربرد مدل‌های آمیخته با پاسخ‌های کرانگینی برای داده‌های واقعی ارائه شده است. در این مثال اثر متغیر زمان بر کرانگین داده‌های ازن تهران با در نظرگرفتن اسپلاین هموار برای پارامتر مکانی توزیع مینیماهای ماهانهٔ ازن بررسی می‌شود.

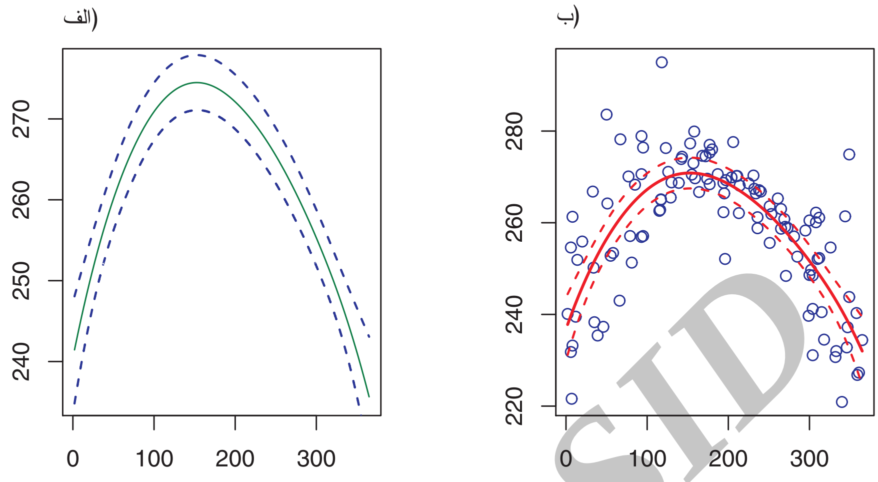
ازن (O_3) در اثر فرآیندهای فتوشیمیایی متفاوت در جو حاصل می‌شود. این فرآیندها به شدت تحت تاثیر شرایط اقلیمی حاکم بر منطقهٔ جغرافیایی هستند. بعنوان مثال انتظار می‌رود ازن در یک روز آفتابی بدون باد در تابستان دارای غلظت بیشتری باشد(اسمیت، ۲۰۰۴). ازن بصورت ناهمگن در لایه‌های استراتوسفر و تروپوسفر جو پراکنده شده است. از آنجایی 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 ماهواره‌ای ازن کلی که بیشتر مقادیر آن را ازن استراتوسفری تشکیل داده، برای استخراج مینیماهای ماهانه و سپس برازش توزیع GEV به این مقادیر با در نظر گرفتن اسپلاین هموار برای پارامتر مکانی توزیع استفاده می‌شود. داده‌های ماهواره‌ای ازن کلی شهر تهران بصورت روزانه از سال ۱۳۷۵ لغایت ۱۳۸۴ با واحد دابسن در سایت ناسا [5] ثبت گردیده است. 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 ماهانهٔ این داده‌ها در نظر گرفته می‌شود. همچنین وجود داده‌های گمشده موجب شد که در مجموع مینیماهای ماهانه ۱۱۴ ماه استخراج شود. بنابراین برای بررسی تغییرات ازن مفید جو و تعیین روند آن در حضور متغیر تبیینی زمان مینیماهای ماهانه ازن کلی شهر تهران با استفاده از مدل آمیختهٔ خطی و اسپلاین هموار با $K = ۱۰$ گره بصورت

$$y_i|\boldsymbol{\beta}, \boldsymbol{u}, \sigma, \xi \sim GEV(\mu_i, \sigma, \xi)$$

$$\mu_i = \beta_۰ + \beta_۱ Day_i + \sum_{k=۱}^{۱۰} u_k |Day_i - k_k|^۳, \quad i = ۱, \ldots, ۱۱۴$$

[5] http://toms.gsfc.nasa.gov/eptoms/ep_v8

Archive of SID
www.SID.ir



شکلِ ۱: (الف) نمودار برآورد تابع اسپلاین هموار، (ب) نمودار پراکنش مینیماهای ماهانه ازن، منحنی برازش شده و نوار اطمینان ۹۵٪ در مقابل متغیر تبیینی روز

مدل‌بندی می‌شوند. برای برازش مدل به مشاهدات توزیع‌های پیشین زیر انتخاب شده‌اند:

$$\beta_0, \beta_1 \sim N(0, 10^4), \quad u_k \sim N(0, \sigma_u^2)$$
$$\sigma \sim IG(0/1, 0/1), \quad \xi \sim U(-0/5, 0/5)$$

مدل فوق با پیشین گامای معکوس برای σ_u^2 براساس نمونه‌ای به حجم ۵۰۰۰ از زنجیر شبیه‌سازی شده با تکرار ۱۵۰۰۰۰، مرحلهٔ داغیدن ۵۰۰۰۰ و تاخیر بیستم به مشاهدات برازش گردید. تاخیر در همگرایی زنجیر و ناسره شدن توزیع پسین سایر پارامترها برای پیشین گامای معکوس σ_u^2 موجب در نظر گرفتن پیشین $U(0, 10)$ با اثر مناسب بر آمیختگی زنجیر شد. در نتیجه با تکرار، مرحلهٔ داغیدن و تاخیر یکسان تحلیل بیزی با پیشین یکنواخت برای σ_u^2 تکرار گردید. تحلیل بیزی مینیماهای ماهانهٔ ازن شهر تهران با توجه به شکل ۱ بیانگر چند نکته است. پارامتر مکانی توزیع GEV مینیماهای ازن تهران در شکل ۱ـالف روند افزایشی و سپس کاهشی را در برابر زمان نشان می‌دهد. شکل ۱ـب نمودار پراکنش مینیماهای ماهانهٔ ازن را نشان می‌دهد، در این نمودار خطپر منحنی مدل برآورد شده بوسیلهٔ میانگین توزیع پسین برازش مناسبی به مشاهدات و خط‌چین نوار اطمینان ۹۵٪ بیزی را نشان می‌دهد. همانطور که ملاحظه می‌شود مینیماهای ماهانهٔ ازن با روند افزایشی همراه است و این روند در تابستان به بیشترین مقدار می‌رسد. با گذر از تابستان، مینیماهای ماهانهٔ ازن با روند کاهشی در روزهای پایانی فصل به کمترین مقدار خود می‌رسد.

Archive of SID
www.SID.ir

به منظور بررسی مناسبت مدل اسپلاین هموار در مقایسه با مدل رگرسیونی خطی و مدل ثابت از ملاک اطلاع انحراف[6] (DIC) (اسپیگ‌هالتر و همکاران، ۲۰۰۲) استفاده شده است. تحلیل بیزی با شرایط یکسان برای دو مدل ثابت $GEV(\mu, \sigma, \xi)$ و خطی $GEV(\mu_i = \beta_0 + \beta_1 x_i, \sigma, \xi)$ تکرار و از بین سه مدل ثابت، خطی و هموارساز مدل دارای کوچکترین DIC انتخاب می‌شود. همانطور که ملاحظه می‌شود مدل آمیخته با اسپلاین هموار دارای کوچکترین مقدار DIC در بین سه مدل می‌باشد.

DIC	مدل
۳۰۳۷	ثابت
۳۰۲۹	خطی
۲۹۵۲	اسپلاین هموار

جدول ۱. مقادیر ملاک اطلاع انحراف.

۵ بحث و نتیجه‌گیری

در این مقاله مدل‌های آمیختهٔ خطی تعمیم‌یافته با استفاده از اسپلاین هموار برای تحلیل داده‌های ازن با رهیافت بیزی بکار گرفته شد. بیان اسپلاین هموار با مدل آمیخته و انجام استنباط با رهیافت بیزی برای اثرات متغیرهای تبیینی پیچیده بر پاسخ‌های کرانگینی براحتی انجام گرفت. در صورتی که در رهیافت بسامدی نیاز به روشی برای برآورد پارامتر هموار و تعیین درستی آن است. همچنین موجود نبودن برآوردگرهای درستنمایی به ازای بعضی از مقادیر پارامتر شکل انگیزه‌ای برای استفاده از رهیافت بیزی در این مقاله را فراهم نمود. تعمیم مدل به پارامتر مقیاس توزیع و بررسی مناسبت آن، بررسی توزیع‌های پیشین دیگر، تحلیل حساسیت مدل به این پیشین‌ها و بکارگیری مولفهٔ زمان و فضایی در کنار بررسی اثرات هموار متغیرهای تبیینی با در نظر گرفتن آنها به شکل مدل‌های آمیختهٔ تعمیم‌یافته از جمله مسائل مهم دیگری است که نیاز به مطالعه بیشتر دارد.

[6] Deviance information criteria

Archive of SID
www.SID.ir

مراجع

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

[2] 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.

[3] 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.

[4] 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.

[5] Fahrmeir, L. and Lang, S. (2001). Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors, *Applied Statistics*, **50**, 201-220.

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

[7] Lambert, P. C., Sutton, A. S., Burton, P. R., Abrams, K. R. and Jones, D. R. (2005). How Vague is Vague? A Simulation Study of the Impact of the Vague Prior Distributions in MCMC Using WinBUGS, *Statistics in Medicine*, **24**, 2401-2428.

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

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

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

[11] Ruppert, D. (2002). Selecting the Number of Knots for Penalized

Archive of SID
www.SID.ir

Splines, *Journal of Computational and Graphical Statistics*, **11**, 735-757.

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

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

[14] 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 & Hall CRC Press, London.

[15] Spiegelhalter, D., Best, N. and carlin, B. (2002). Bayesian Measure of Model Complexity and Fit, *Journal of the Royal Statistical Society, Series B*, **64**, 583-639.

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

Archive of SID

www.SID.ir