

مدل بندی داده‌های بقا طولانی - مدت با استفاده از تابع مفصل کلایتون

مرضیه روزبهانی* ، دانشجوی کارشناسی ارشد آمار، دانشگاه شهید چمران اهواز، m-ruzbahani@mscstu.scu.ac.ir

محمد رضا آخوند، عضو هیأت علمی گروه آمار، دانشگاه شهید چمران اهواز، mra.biostat@gmail.com

سلیمان خیری، عضو هیأت علمی گروه آمار و اپیدومولوژی، دانشگاه علوم پزشکی شهرکرد، kheiri@hbi.ir.ac.ir

چکیده: در حالت کلی داده‌های بقا چند متغیره به هم وابسته هستند و یکی از راه‌های در نظر گرفتن وابستگی بین داده‌ها استفاده از تابع مفصل می‌باشد. در این مقاله قصد داریم داده‌های بقا طولانی مدت را با استفاده از مفصل کلایتون مدل‌سازی کنیم. برای این منظور برآورد پارامترها را به کمک روش بیزی به دست می‌آوریم. از آنجایی که توزیع‌های پسین دارای فرم بسته‌ای نمی‌باشند، برآورد مشخصات توزیع‌های پسین را با به کارگیری روش مونت کارلوی زنجیر مارکوفی به دست خواهیم آورد.

کلمات کلیدی: تابع مفصل، بقا طولانی مدت، روش بیزی، روش‌های مونت کارلوی زنجیر مارکوفی

مقدمه

مدل شکنندگی است که یک یا چند اثر تصادفی به منظور نشان دادن وابستگی بین مشاهدات در مدل وارد می‌شود و زمان‌های بقا به طور شرطی با در نظر گرفتن اثر تصادفی از هم مستقل در نظر گرفته می‌شوند. واپیل و همکاران [۴] برای اولین بار مدل‌های شکنندگی را مطرح کردند. در سال‌های اخیر، توابع مفصل به عنوان یک ابزار مناسب برای بیان وابستگی بین متغیرها مطرح شده‌اند و در بیشتر زمینه‌های پزشکی، مالی و بیمه مورد استفاده قرار گرفته‌اند. مفصل‌ها توابعی هستند که تابع توزیع چند متغیره را به تابع توزیع حاشیه‌ای یک متغیره متصل می‌کنند.

در آنالیز بقا به طور کلی فرض می‌شود که همه افراد در معرض خطر هستند و پیشامد مورد نظر را تا

در بسیاری از زمینه‌های کاربردی با داده‌های زمان شکستی برخورد می‌کنیم که ساختاری دو یا چند متغیره دارند، در این حالت نمی‌توان زمان‌های شکست را مستقل فرض نمود. داده‌های بقا چند متغیره در حالت کلی به هم وابسته هستند و مطالعه‌ای از وابستگی بین متغیرها معمولاً مورد توجه محققین است. از جمله افرادی که بر روی مدل‌سازی داده‌های بقا چند متغیره کار کرده‌اند می‌توان به اسلانید و همکاران [۱]، رومئو و همکاران [۲] و هنگل [۳] اشاره نمود. برای در نظر گرفتن ساختار وابستگی میان چنین داده‌هایی روش‌های مختلفی مورد استفاده قرار می‌گیرد. که رایج‌ترین آن



$$\tau_{\alpha}(T_1, T_2) = \frac{\alpha}{\alpha+2}$$

تابع بقا شفایافته

در مدل‌های شفایافته آمیخته فرض بر این است که جامعه به صورت ناهمگن به دو زیر جامعه از افراد مصون یا شفایافته و افراد در معرض خطر تقسیم شده است. در این مدل‌ها فرض می‌شود که هر فرد با احتمال p شفایافته و با احتمال $1-p$ در معرض خطر است. در مدل‌های آمیخته تابع بقا برای کل افراد جامعه به صورت زیر است:

$$S_j(t_j) = p_j + (1 - p_j) S_o(t_j)$$

که p درصدی از افراد شفایافته است و $S_o(t_j)$ تابع بقا برای افراد شفایافته است. که معمولاً برای آن توزیع نمایی، وایبل و گامبرتر در نظر گرفته می‌شود.

تشکیل تابع درستنمایی

فرض کنید که زمان‌های طول عمر (T_{i1}, T_{i2}) و زمان سانسور شدن (C_{i1}, C_{i2}) برای $i = 1, \dots, n$ از یکدیگر مستقل هستند. در نتیجه $t_{ij} = \min(T_{ij}, C_{ij})$ طول عمر مشاهده شده برای فرد i ام و $\delta_{ij} = I(t_{ij} = T_{ij})$ نشانگر سانسور برای $j = 1, 2$ است. در نتیجه تابع درستنمایی برای داده‌ها به صورت زیر به دست می‌آید:

$$L(\theta) = \prod_{i=1}^n f(t_{i1}, t_{i2})^{\delta_{i1}\delta_{i2}} \frac{\partial S(t_{i1}, t_{i2}|\theta)^{\delta_{i1}(1-\delta_{i2})}}{\partial t_{i1}} \times \frac{\partial S(t_{i1}, t_{i2}|\theta)^{\delta_{i2}(1-\delta_{i1})}}{\partial t_{i2}} \times S(t_{i1}, t_{i2})^{(1-\delta_{i1})(1-\delta_{i2})}$$

با در نظر گرفتن توزیع وایبل برای حاشیه‌ها $T_j \sim weibull(r_j, \lambda_j)$ متغیرهای کمکی از طریق پارامتر $\lambda_{ij} = (\beta_j + \beta_{ij} X_{ij})$ وارد مدل می‌شوند و تابع درستنمایی با در نظر گرفتن فرضیات بالا به صورت

پایان مطالعه تجربه می‌کنند. اگر قسمتی از جامعه پیشامد مورد علاقه را تجربه نکنند این‌گونه از افراد تحت عنوان افراد با بقا طولانی مدت یا شفایافته نامیده می‌شوند. رایج‌ترین نوع از مدل‌های شفایافته، مدل آمیخته است که توسط بوگ [۵] ارائه شده است. لئوزادا و همکاران [۶] از تابع مفصل FGM برای مدل‌بندی داده‌های بقا دو متغیره استفاده کردند. خیری و همکاران [۷] از مدل‌های شکنندگی برای مدل‌بندی داده‌های قوز قرنیه استفاده کردند. در این مقاله قصد داریم به مدل‌سازی داده‌های بقا طولانی مدت با استفاده از توابع مفصل پردازیم و در پایان از این مدل‌ها برای مدسازی داده‌های پیوند قوز قرنیه استفاده کنیم.

مدل آماری

مفصل‌ها توابع چند متغیره‌ای هستند که توابع توزیع حاشیه‌ای آن‌ها به صورت یکنواخت روی بازه‌ی $(0, 1)$ توزیع شده‌اند. در حالت دو متغیره فرض کنید که (T_1, T_2) متغیرهای تصادفی پیوسته با توابع بقا حاشیه‌ای به ترتیب (S_1, S_2) باشند. در این صورت تابع بقا توام را با استفاده از مفصل C_{ϕ} می‌توان به صورت زیر نوشت:

$$S(t_1, t_2) = C_{\phi}(S_1(t_1), S_2(t_2))$$

که با استفاده از تابع مفصل کلایتون تابع بقا توام به صورت زیر به دست می‌آید:

$$S(t_1, t_2) = (S_1(t_1)^{-\alpha} + S_2(t_2)^{-\alpha} - 1)^{-\frac{1}{\alpha}}$$

که در آن $S_1(t_1)$ و $S_2(t_2)$ توابع بقا حاشیه‌ای برای T_1 و T_2 به ترتیب α پارامتر وابستگی است. اگر $\alpha \rightarrow 0$ حاشیه‌ها مستقل خواهند بود یعنی $S(t_1, t_2) = S(t_1) \times S(t_2)$. با مشتق‌گیری از تابع بقا توام نسبت به t_1 و t_2 می‌توان تابع چگالی توام را به دست آورد. با استفاده از ضریب همبستگی تاوکندال می‌توان میزان همبستگی بین متغیرها را بهتر درک کرد که در تابع مفصل کلایتون از فرمول زیر به دست می‌آید:



به دلیل پیچیدگی و ابعاد گسترده توزیع پسین توام به دست آمده، امکان محاسبه توزیع پسین پارامترها به روش تحلیلی وجود ندارد. برای تقریب توزیع پسین پارامترها از روش مونت کارلوی زنجیر مارکوفی استفاده می‌گردد. در این روش نمونه‌هایی تصادفی از توزیع پسین تولید می‌شود و بر اساس نمونه‌ها استنباط در مورد پارامترها صورت می‌گیرد.

زیر به دست می‌آید

$$L = \prod_{i=1}^n (\alpha + 1)^{\delta_{i1}\delta_{i2}} \left(f_{i1}(t_{i1}) S_{i1}(t_{i1})^{-\alpha-1} \right)^{\delta_{i1}} \left(f_{i2}(t_{i2}) S_{i2}(t_{i2})^{-\alpha-1} \right)^{\delta_{i2}} \left(S_{i1}(t_{i1})^{-\alpha} + S_{i2}(t_{i2})^{-\alpha} - 1 \right)^{-\delta_{i1}-\delta_{i2}-\frac{1}{\alpha}}$$

که

$$S_j(t_j) = p_j + (1 - p_j) \exp(-\lambda_j t_j^{r_j})$$

و

$$f_j(t_j) = (1 - p_j) \lambda_j r_j t_j^{r_j-1} \exp(-\lambda_j t_j^{r_j})$$

به ترتیب تابع بقا شفا یافته آمیخته وایل و چگالی آمیخته وایل برای $j = 1, 2$ است.

مثال کاربردی

از آنجایی که مهمترین علت شکست پیوند، دفع عضو عضو پیوندی می‌باشد. بررسی و تعیین دقیق عوامل موثر بر دفع پیوند قرنيه از اهمیت ویژه‌ای برخوردار است. در این مقاله اطلاعات ۱۱۹ بیمار که طی سال‌های ۸۰-۶۵ تحت عمل پیوند قرنيه دو طرفه قرار گرفته‌اند، مورد تجزیه و تحلیل قرار گرفته است. متغیرهای کمکی موجود در این داده‌ها عبارتند از: جنس، سن، قطر قرنيه‌دهنده، قطر بستر گیرنده، تازگی قرنيه، ورم ملتحمه بهاره، واسکولاریزیشن و پیوند مجدد هستند که در مدل وارد شده‌اند.

برای انجام تحلیل بیزی از توزیع پیشین ناآگاهی بخش برای پارامترها به صورت $r_j \sim \beta_{ij} \sim N(0, 100^2)$ ، $\alpha \sim p_j \sim \text{Beta}(1, 1)$ ، $\text{Gamma}(0/1, 0/001)$ استفاده گردید.

برای انجام این تحلیل از نرم افزار OpenBugs استفاده شده است و نتایج حاصل براساس برآوردهای پسین مدل، که شامل میانگین، انحراف معیار و فاصله باورمند ۹۵ درصد است در جدول ۱ آمده است. تمام شبیه سازی‌ها براساس دو زنجیر به طول ۵۰ هزار که ۱۰ هزار نمونه اول به عنوان دوره داغیدن در نظر گرفته شده است و از هر ۵ مشاهده یک مشاهده انتخاب شده است بدست آمده است.

چگالی پیشین و پسین

پس از مدل سازی داده‌های بقا طولانی مدت می‌توان از روش‌های مختلفی برای برآورد پارامترها استفاده نمود، که از جمله آنها می‌توان روش‌های بیزی را نام برد. از مزیت‌های روش بیزی نسبت به روش کلاسیک این است که در صورتی که اطلاعات اضافی در مورد پارامترها در اختیار باشد می‌توان آن را از طریق توزیع پیشین وارد مدل نمود و در صورتی که چنین اطلاعاتی وجود نداشته باشد از توزیع پیشین ناآگاهی بخش که هیچ‌گونه اطلاعات اضافی وارد مدل نمی‌کنند، می‌توان استفاده کرد. با در نظر گرفتن توزیع‌های پیشین برای پارامترهای مدل به صورت $r_j \sim \text{Gamma}(a_j, b_j)$ ، $\alpha \sim p_j \sim \text{Beta}(v_j, u_j)$ ، $\beta_{ij} \sim N(\mu_{ij}, \sigma_{ij})$ ، $\text{Gamma}(c, d)$ توزیع پسین توام پارامترها به صورت زیر به دست می‌آید:

$$\pi(\alpha, r_1, r_2, \beta_{11}, \beta_{12} | X, \delta_1, \delta_2) \propto \prod_{i=1}^n (\alpha + 1)^{\delta_{i1}\delta_{i2}} \left((1 - p_1) e^{(\beta_{01} + \beta_{k1}) t_1^{r_1-1}} \left(p_1 + (1 - p_1) e^{-\exp(\beta_{01} + \beta_{k1}) t_1^{r_1}} \right)^{-\alpha-1} \right)^{\delta_{i1}} \left((1 - p_2) e^{(\beta_{02} + \beta_{k2}) t_2^{r_2-1}} \left(p_2 + (1 - p_2) e^{-\exp(\beta_{02} + \beta_{k2}) t_2^{r_2}} \right)^{-\alpha-1} \right)^{\delta_{i2}} \left(\left(p_1 + (1 - p_1) e^{-\lambda_1 t_1^{r_1}} \right)^{-\alpha} + \left(p_2 + (1 - p_2) e^{-\lambda_2 t_2^{r_2}} \right)^{-\alpha} - 1 \right)^{-\left(\frac{1}{\alpha} + \delta_{i1} + \delta_{i2} \right)} e^{a_1 r_1 a_1 - 1 - b_1 r_1 b_1 r_1 a_1 - 1 - b_2 r_2 r_2 c d a c - 1 - d \alpha p_2 v_2 - 1 - (1 - p_2) u_2 - 1} \left(-\frac{1}{r_1 \sigma_{k1}^2} (\beta_{01} - \mu_{01})^2 \right) e \left(-\frac{1}{r_2 \sigma_{k2}^2} (\beta_{02} - \mu_{02})^2 \right) \left(-\frac{1}{r_2 \sigma_{k1}^2} (\beta_{k1} - \mu_{k1})^2 \right) e \left(-\frac{1}{r_2 \sigma_{k2}^2} (\beta_{k2} - \mu_{k2})^2 \right)$$

Bayesian approach based on copulas," Lifetime Data Anal, 12: 205-222.

- [3] D. Hanagal.(2011). " Modeling survival data using frailty models," Chapman and Hall, Boca Raton.
- [4] J. W. Vaupel and K. G. Manton.(1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality," Demography, 16, 439-454.
- [5] J. W. Boag.(1949). "Maximum likelihood estimates of the proportion of patients cured by cancer therapy," Journal of the Royal Statistical Society, 11, 15-53
- [6] F.Louzada, A. K. Suzuki. and V. G. Cancho. (2013). "The FGM long-term bivariate survival copula model: Modeling bayesian estimation and case influence diagnostics," Communications in Statistics – Theory and Methods, 42:4,673-691.
- [7] S. Kheiri, A. Kimber and M. R. Meshkani(2007). "Bayesian analysis of an inverse Gaussian correlated frailty model," Computational Statistics and Data Analysis, 51, 5317 – 5326.
- [8] M. Rahimzadeh, E. Hajizadeh and F. Eskandari(2011). "Non-mixture cure correlated frailty models in Bayesian approach," Journal of Applied Statistics, 38:8, 1651-1663.

نتایج

تحلیل چندگانه داده‌ها با استفاده از مفصل کلایتون نشان داد که سن بیمار در زمان پیوند، قطر قرنیه دهنده، قطر بستر گیرنده، تازگی قرنیه، ورم ملتحمه بهاره و واسکولاریزیشن عوامل معنی‌داری در زمان دفع پیوند بوده‌اند. اما عواملی چون جنس بیمار و پیوند مجدد تاثیری بر زمان بقای پیوند نداشته‌اند. رحیم زاده و همکاران [۸] در مطالعه خود که با استفاده از مدل‌های شکنندگی انجام شده است نشان دادند که متغیرهای سن و واسکولاریزیشن دارای اثری معنی‌دار بر روی دفع پیوند می‌باشد.

جدول ۱: میانگین پسین، انحراف استاندارد و بازه باورمند ۰/۹۵ برای پیوند قوز قرنیه

پارامتر	میانگین	انحراف استاندارد	بازه باورمند ۰/۹۵
جنس	-0.369	0.583	(-1.539 , 0.775)
سن	0.077	0.032	(0.011 , 0.137)
قطر قرنیه دهنده	-3.244	0.830	(-4.489 , -1.67)
قطر بستر گیرنده	4.119	1.493	(2.274 , 6.675)
تازگی قرنیه	1.265	0.645	(0.052 , 2.562)
ورم ملتحمه بهاره	3.049	0.874	(1.259 , 4.652)
واسکولاریزیشن	3.36	0.758	(1.824 , 4.793)
پیوند مجدد	-2.302	1.525	(-5.664 , 0.381)
نسب شفایافتگی	0.545	0.073	(0.39 , 0.679)
میزان همبستگی	0.498	0.45	(0.016 , 1.649)

مراجع

- [1] H. Aslanidou, D. K. Dey, and D. sinha. (2008). "Bayesian analysis of multivariate survival data using monte carlo methods," Canadian Journal of statistics. 1, 33-48
- [2] J. S. Romeo, N. I. Tanaka and A. C. Pedroso de Lima.(2006). "Bivariate survival odeling a