

## توزیع فوق هندسی تعمیم یافته دوجمله‌ای منفی در مدل شفایافته زمان پیشرفت

میترا رحیم‌زاده<sup>۱</sup>، احمدرضا باغستانی<sup>۲</sup>، بهروز کاوه‌ئی<sup>۳</sup>

<sup>۱</sup> مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشگاه علوم پزشکی البرز

<sup>۲</sup> دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی

<sup>۳</sup> سازمان سنجش آموزش کشور

تاریخ دریافت: ۱۳۹۱/۸/۲۸ تاریخ آخرین بازنگری: ۱۳۹۲/۵/۱۳

**چکیده:** در تحلیل داده‌های بقا، اگر در پایان مطالعه با درصد بالایی از سانسور مواجه شویم، چنانچه طول مدت مطالعه به اندازه کافی طولانی باشد، بهتر است از مدل‌های شفایافته استفاده شود. این مدل‌ها با ارائه فرایندی که بر اساس توزیع متغیر پنهان استوار است، در دهه اخیر مورد توجه قرار گرفته است. در این مقاله با در نظر گرفتن توزیع فوق هندسی تعمیم‌یافته دوجمله‌ای منفی برای متغیر پنهان مدل جدیدی برای تحلیل داده‌های بقا طولانی مدت به دست می‌آید. بر آورد بیزی پارامترهای مدل با روش‌های عددی زنجیر مارکوفی مونت کارلویی به دست آورده می‌شود. کاربرد مدل برای داده‌های کارآزمایی بالینی درمان بیماران مبتلا به سیروز کبدی و داده‌های شبیه‌سازی شده، نشان داده می‌شود. با معیار اطلاع انحرافی، مدل فوق هندسی تعمیم‌یافته دوجمله‌ای منفی برآزش بهتری به داده‌ها را نشان می‌دهد.

---

آدرس الکترونیک مسئول مقاله: احمدرضا باغستانی، baghestani.ar@gmail.com  
کد موضوع بندی ریاضی (۲۰۱۰): ۶۲N۰۱، ۶۲F۱۵

واژه های کلیدی : بقا طولانی مدت، مدل شفایافته زمان پیشرفت، توزیع فوق هندسی تعمیم یافته دو جمله ای منفی.

## ۱ مقدمه

در تحلیل داده های بقا طولانی مدت، با توجه به پیشرفت های به دست آمده در درمان بیمارها، در انتهای مطالعه با تعداد زیادی از موارد سانسور شده مواجه خواهیم شد. اگر طول مدت مطالعه به اندازه کافی طولانی باشد، استفاده از مدل های نسبت شفایافتگی<sup>۱</sup> نتایج بهتری خواهد داشت. دو رده عمده از این مدل ها تحت عنوان مدل آمیخته و مدل نا آمیخته در تحلیل داده های بقا طولانی مدت مطرح شده اند. در مدل شفایافته آمیخته (بوگ، ۱۹۴۹) هر فردی از جامعه با احتمال  $p$  دارای بقا یک و با احتمال  $1-p$  دارای بقایی کمتر از یک است، که به کمک یکی از توابع متداول بقا برآورد می شود (مالر و ژو، ۱۹۹۶). تابع بقا برای کل افراد مورد مطالعه در این مدل به صورت

$$S_p(t, x) = p(x) + (1 - p(x))S(t, x), \quad S(\infty) = 0, \quad S_p(\infty) = p > 0$$

است، که در آن  $p(x)$  نسبت شفایافتگی است که برای برآورد آن می توان از تابع ربط لوژستیک یا لگ-لگ مکمل استفاده کرد و  $S(t, x)$  تابع بقا برای افراد در معرض رخدادهای پیشامد می باشد. مدل شفایافته نا آمیخته که توسط یا کولوف و سودیکوف (۱۹۹۶) ارائه شد و سپس چن و همکاران (۱۹۹۹) مکانیزمی بر اساس توزیع متغیر پنهان برای آن ارائه کرد، که این فرایند نیز توسط کونر و همکاران (۲۰۰۷) مورد بسط قرار گرفت. این مدل بر اساس توزیع متغیر پنهان  $N$  (تعداد سلول های سرطانی فعال پس از درمان بیماری) استوار است. اگر در درمان بیماری سرطان زمان لازم برای تولید تومور قابل شناسایی با  $Z_i, i = 1, \dots, N$  نشان داده شود و  $Z_i$  هامتغیرهای تصادفی مستقل و هم توزیع با تابع توزیع تجمعی  $F(\cdot)$  و مستقل از  $N$  باشند در این صورت زمان عود مجدد بیماری بر اساس اولین فرایند فعال سازی کونر به صورت  $T = \{\min Z_i, 1 \leq i \leq N\}$  به دست می آید، که  $P(Z_i = \infty) = 1$

<sup>۱</sup> Cure rate models

است. در نتیجه در صورتی که متغیر پنهان  $N$  دارای توزیع پواسون با پارامتر  $\theta$  باشد، تابع بقا برای متغیر تصادفی زمان  $T$  به صورت  $S_p(t; x) = \exp(-\theta(x)F(t))$  به دست می‌آید، که در آن  $F(t)$  یک تابع توزیع تجمعی برای متغیر نامنفی است و مقدار  $\theta(x)$  از روی متغیرهای کمکی و با تابع ربط نمایی برآورد می‌شود.

در سال‌های اخیر توزیع‌های متفاوتی برای متغیر پنهان  $N$  در نظر گرفته شده است. برای مثال، اگر توزیع متغیر پنهان  $N$  برنولی در نظر گرفته شود مدل شفایافته آمیخته حاصل می‌شود. همچنین کونر و همکاران (۲۰۰۷) از توزیع هندسی، برنولی و دوجمله‌ای، بورگر و همکاران (۲۰۱۲) توزیع سری‌های توانی، رودریگز و همکاران (۲۰۰۹) توزیع پواسون کانوی-ماکسول<sup>۲</sup> و باغستانی و همکاران (۲۰۱۳) توزیع تعمیم‌یافته پواسون لیندلی را در مدل‌های شفایافته زمان پیشرفت ارائه کردند. شایان ذکر است توزیع متغیر پنهان می‌تواند هر یک از توزیع‌های گسسته که دارای جرم احتمال در نقطه صفر هستند، به منظور تبیین احتمال شفایافتگی، باشد. به نظر می‌رسد رابطه نزدیکی بین توزیع‌هایی که در تحلیل داده‌های شمارشی و مدل‌های شفایافته زمان پیشرفت مورد استفاده قرار می‌گیرد، وجود دارد. مشکل دیگری که علاوه بر بیش پراکنشی در تحلیل داده‌های بقا وجود دارد، چوله بودن داده‌ها است، که دلیلی دیگر برای ارائه توزیع‌های متفاوت برای حل این مشکل است.

در این مقاله مدلی ارائه شده است، که در آن توزیع متغیر پنهان فوق هندسی تعمیم‌یافته دوجمله‌ای منفی<sup>۳</sup> (HGNB) است و حاصل آن مدلی انعطاف پذیر است که تعمیم‌یافته مدل‌های دوجمله‌ای منفی، لیندلی پواسون و تعمیم‌یافته آمیخته هندسی می‌باشد. در بخش بعدی توزیع فوق هندسی تعمیم‌یافته دوجمله‌ای منفی معرفی می‌شود. در بخش ۳ مدل شفایافته زمان پیشرفت با توزیع فوق هندسی تعمیم‌یافته دوجمله‌ای منفی و در بخش ۴ برآورد پارامترهای مدل با رهیافت بی‌زی ارائه و با معیار اطلاع انحرافی<sup>۴</sup> ( $DIC$ ) ارزیابی می‌شوند. در بخش ۵ کاربرد مدل در یک سری داده واقعی ارائه خواهد شد و در بخش ۶ چگونگی شبیه‌سازی مدل

<sup>۲</sup> Conway-Maxwell-Poisson distribution

<sup>۳</sup> Hypergeometric generalized negative binomial distribution

<sup>۴</sup> Deviance Information Criteria

ارائه شده نشان داده می شود. بحث و نتیجه گیری در بخش نهایی ارائه می شود.

## ۲ توزیع فوق هندسی تعمیم یافته دو جمله ای منفی

در تحلیل داده های شمارشی با مشکل بیش پراکنشی، توزیع دو جمله ای منفی جایگزین مناسبی برای توزیع پواسون است. این توزیع، نمونه مشهوری از ترکیب توزیع گسسته پواسون با توزیع پیوسته گاما است. مثال دیگر که توسط قیتانی و همکاران (۲۰۰۲) ارائه شده است توزیع فوق هندسی تعمیم یافته دو جمله ای منفی نامیده می شود که توزیعی آمیخته از توزیع پواسون با توزیع فوق هندسی همشاره است.

فرض کنید  $X$  دارای توزیع پواسون با تابع جرم احتمال

$$f(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots, \lambda > 0$$

باشد. اگر  $\lambda$  دارای تابع چگالی احتمال به صورت

$$g(\lambda) = \frac{\alpha^\nu (\alpha + 1)^{p-\nu}}{\Gamma(p)} \lambda^{p-1} e^{-(\alpha+1)\lambda} {}_1F_1(\nu; p; \lambda), \quad \lambda > 0, \nu, p, \alpha > 0$$

باشد، که در آن  ${}_1F_1(\nu; p; \lambda) = \sum_{n=0}^{\infty} \frac{(\nu)_n \lambda^n}{(p)_n n!}$  تابع فوق هندسی همشاره است و

$$(\nu)_0 = 1, \quad (\nu)_n = \Gamma(\nu + n) / \Gamma(\nu), \quad n = 1, \dots$$

آنگاه تابع چگالی کناری  $X$  به صورت

$$f(x) = \frac{\alpha^\nu (\alpha + 1)^{p-\nu} (p)_x}{x! (\alpha + 2)^{x+p}} {}_2F_1(\alpha, x + p; p; \frac{1}{\alpha + 2}), \quad x = 0, 1, \dots \quad (1)$$

به دست می آید، که در آن  ${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n z^n}{(c)_n n!}$  تابع فوق هندسی است. توزیع فوق هندسی تعمیم یافته دو جمله ای منفی با نماد  $HGNB(\alpha, \lambda, p)$  نشان داده می شود. این توزیع در حالت های خاص به توزیع دو جمله ای منفی، لیندلی پواسون و تعمیم یافته آمیخته هندسی تقلیل می یابد.

<sup>5</sup> Confluent hypergeometric function

تابع مولد احتمال این توزیع به صورت

$$G_x(t) = \frac{[1 - (t-1)/(\alpha+1)]^{\lambda-p}}{[1 - (t-1)/\alpha]^\lambda}, \quad 0 < t < 1$$

است، میانگین، واریانس و شاخص بیش پراکندگی این توزیع نیز به صورت

$$E(X) = \frac{\lambda + p\alpha}{\alpha(\alpha + 1)}, \quad V(X) = \frac{\lambda(\alpha + 1)^2 - (\lambda - p)\alpha^2(\alpha + 2)}{\alpha^2(\alpha + 1)^2}$$

$$ID = \frac{\alpha(\alpha + 1)^2 - (\lambda - p)\alpha^2(\alpha + 2)}{\alpha(\alpha + 1)(\lambda + p\alpha)} > 1$$

به دست می آید. بنابراین این توزیع برای تمام مقادیر  $\lambda$  و  $p$  دارای بیش پراکنش است. همچنین بیش پراکنشی برای توزیع دو جمله ای منفی، لیندلی پواسون و تعمیم یافته آمیخته هندسی را وقتی  $(\lambda, p)$  به ترتیب برابر با  $(\lambda, \lambda)$ ،  $(2, 1)$  و  $(1, 2)$  باشد، می توان به دست آورد (قیتانی و همکاران، ۲۰۰۲).

### ۳ مدل شفایافته زمان پیشرفت با توزیع فوق هندسی تعمیم یافته دو جمله ای منفی

در مدل شفایافته زمان پیشرفت با فرآیند فعال سازی انعطاف پذیری که توسط کونر و همکاران (۲۰۰۷) ارائه شده است، تابع بقا برای کل جامعه بر اساس اولین فرآیند فعال سازی به صورت

$$S_p = P(N = 0) + \sum_{n=1}^{\infty} P(Z_1 > t, \dots, Z_n > t)P(N = n)$$

$$= \sum_{n=0}^{\infty} S(t)^n P(N = n) = G_N(S(t))$$

تعریف می شود، که در آن  $G_N(\cdot)$  تابع مولد احتمال متغیر تصادفی پنهان  $N$  می باشد.

بنابراین تابع بقا و تابع چگالی برای کل جامعه به صورت

$$S_p(t) = \frac{[(2 + \alpha - S(t))/\alpha + 1]^{\lambda-p}}{[(1 + \alpha - S(t))/\alpha]^\lambda}$$

$$f_p(t) = -f(t)(\lambda - p) \left[ \frac{2 + \alpha - S(t)}{\alpha} + 1 \right]^{\lambda-p-1} \left[ \frac{1 + \alpha - S(t)}{\alpha} \right]^{-\lambda}$$

$$+ \lambda f(t) \left[ \frac{2 + \alpha - S(t)}{\alpha} + 1 \right]^{\lambda-p} \left[ \frac{1 + \alpha - S(t)}{\alpha} \right]^{-\lambda-1}$$

۵۰..... توزیع فوق هندسی تعمیم یافته دو جمله ای منفی در مدل شفایافته

به دست می آید، که در آن  $S(t)$  تابع بقا زمان پیشرفت در زمان  $t$  است و می تواند هر یک از توابع متداول بقا مانند وایبول یا نمایی در نظر گرفته شود. نسبت شفایافتگی به صورت  $P(N = 0) = \left(\frac{\alpha}{1+\alpha}\right)^\lambda \left(\frac{\lambda+\alpha}{1+\alpha}\right)^{\lambda-p}$  به دست می آید، که در آن  $\alpha = \exp(\beta'X)$ .

#### ۴ برآورد پارامترهای مدل

فرض کنید  $n$  تعداد افراد شرکت کننده در مطالعه و  $N_i$  تعداد سلول های سرطانی فعال باقی مانده پس از درمان برای فرد  $i$ ام  $i = 1, \dots, n$  باشد. اگر  $Y_i$  و  $C_i$  به ترتیب مشخص کننده زمان بقا و سانسور شدگی برای فرد  $i$ ام باشد زمان مشاهده شده برای این فرد به صورت  $T_i = \min\{Y_i, C_i\}$  است به طوری که برای تابع نشانگر  $\sigma_i$  وقتی  $\sigma_i = 1$  خواهیم داشت  $T_i = Y_i$  و در صورتی که  $\sigma_i = 0$ ، آنگاه  $T_i = C_i$  خواهد بود. بنابراین برای فرد  $i$ ام ماتریس مشاهدات  $D_{obs} = \{T_i, \sigma_i, X_i\}$  را خواهیم داشت که در آن  $X_i$  ماتریس در برگیرنده متغیرهای کمکی خواهد بود.

در این مقاله فرض می شود که  $N_i$  و  $Z_i$  ها مستقل از هم و  $N_i$  دارای توزیع فوق هندسی تعمیم یافته دو جمله ای منفی با تابع چگالی احتمال ارائه شده در (۱) است. به طوری که  $\lambda > 0$ ،  $\alpha > 0$  و  $p > 0$  پارامترهای توزیع هستند. به شرط  $N_i = n_i$  زمان های مورد نیاز برای تولید تومور قابل تشخیص  $Z_i$  ها مستقل از هم و دارای توزیع وایبول ارائه شده در (۲) خواهند بود. بنابراین تابع درستیابی به صورت

$$\begin{aligned} L(\alpha, \theta, \beta, \tau, \nu | D_{obs}) &= \prod_{i=1}^n S_p(t)^{1-\sigma_i} \times f_p(t)^{\sigma_i} \\ &= \left\{ \frac{((\lambda + \alpha - S(t))/\alpha + 1)^{\lambda-p}}{((1 + \alpha - S(t))/\alpha)^\lambda} \right\}^{1-\sigma_i} \\ &\times \left\{ -f(t) \left( \frac{\lambda-p}{\alpha} + 1 \right) \left( \frac{\lambda + \alpha - S(t)}{\alpha + 1} \right)^{\lambda-p-1} \left[ \frac{1 + \alpha - S(t)}{\alpha} \right]^{-\lambda} \right. \\ &\left. + \lambda \frac{f(t)}{\alpha} \left[ \frac{\lambda + \alpha - S(t)}{\alpha} + 1 \right]^{\lambda-p} \left[ \frac{1 + \alpha - S(t)}{\alpha} \right]^{-\lambda-1} \right\}^{\sigma_i} \end{aligned}$$

به دست خواهد آمد، که در آن  $S(t)$  تابع بقا زمان لازم برای تولید تومور قابل تشخیص (زمان پیشرفت بیماری) و دارای توزیع وایبول است.

$$S(t) = \exp(-e^\tau t^\nu), \quad \nu > 0, \quad -\infty < \tau < -\infty \quad (2)$$

برای برآورد بیزی پارامترها توزیع‌های پیشین ناآگاهی بخش در نظر گرفته شده است به طوری که تابع درست‌نمایی اثر غالب‌تری بر روی توزیع‌های پسین داشته باشد. بدون آسیب رساندن به کلیت موضوع می‌توان فرض کرد که توزیع‌های پیشین از هم مستقل هستند. برای ضرایب رگرسیون از توزیع ناآگاهی بخش یکنواخت به صورت  $\Pi(\beta) \propto 1$ ، برای پارامترهای توزیع وایبول  $\tau$  و  $\nu$ ، با توجه به دامنه‌شان به ترتیب از توزیع‌های نرمال و گاما و برای پارامترهای  $\lambda$  و  $p$  نیز از توزیع گاما استفاده شده است. در نتیجه توزیع‌های شرطی کامل برای پارامترهای مدل به صورت

$$\begin{aligned} p(\beta_k | D_{obs}, \beta_{(-k)}, \nu, \tau, p, \lambda) &\propto L(\beta, \nu, \tau, p, \lambda | D_{obs}) \times \pi(\beta_k) \\ p(\nu | D_{obs}, \beta, \tau, p, \lambda) &\propto L(\beta, \nu, \tau, p, \lambda | D_{obs}) \times \pi(\nu) \\ p(\tau | D_{obs}, \beta, \nu, p, \lambda) &\propto L(\beta, \nu, \tau, p, \lambda | D_{obs}) \times \pi(\tau) \\ p(p | D_{obs}, \beta, \nu, \tau, \lambda) &\propto L(\beta, \nu, \tau, p, \lambda | D_{obs}) \times \pi(p) \\ p(\lambda | D_{obs}, \nu, \tau, p, \lambda) &\propto L^m(\beta, \lambda | E, D_{obs}) \times \pi(\lambda) \end{aligned}$$

به دست می‌آیند، که در آن  $\beta_{(-k)}$  بردار  $\beta_k$  است که مولفه  $i$ ام آن حذف شده است.  $\pi(\nu)$ ،  $\pi(\tau)$ ،  $\pi(p)$  و  $\pi(\lambda)$  توزیع‌های پیشینی مورد استفاده در مدل هستند.

به واسطه پیچیدگی و ابعاد زیاد توزیع‌های توام فوق، امکان محاسبه توزیع پسینی پارامترهای مدل به روش تحلیلی وجود ندارد، لذا روش‌های زنجیر مارکوف مونت کارلویی، برای استنباط درباره پارامترهای مدل به کار برده می‌شود. برای این منظور با نمونه‌گیری متوالی از توزیع‌های شرطی کامل پارامترها با استفاده از الگوریتم متروپولیس هستینگز، زنجیرهای مارکوفی خواهیم ساخت که توزیع حدی این زنجیرها تقریب مناسبی از توزیع پسینی پارامترهای مدل می‌باشد.

برای مقایسه بیزی دو مدل پواسون (چن و همکاران، ۱۹۹۹) و دوجمله‌ای منفی (کانچو و همکاران، ۲۰۱۱) با مدل فوق هندسی تعمیم یافته دوجمله‌ای منفی معیار اطلاع انحرافی (اشپیگل هالتر و همکاران، ۲۰۰۲) به کار گرفته می‌شود. این معیار هر

۵۲ ..... توزیع فوق هندسی تعمیم یافته دو جمله ای منفی در مدل شفایافته

دو جنبه برآزش و پیچیدگی مدل را دربر داشته و مشکلات تکنیکی مربوط به پیشین‌های ناآگاهی بخش را نداشته و به صورت  $DIC = \bar{D}(\theta) + P_D$  تعریف می‌شود، که در آن  $\bar{D}(\theta)$  میانگین پسین انحراف بوده و میزان برآزش را نشان می‌دهد و  $P_D$  تعداد پارامترهای موثر است که میزان پیچیدگی مدل را نشان می‌دهد. براساس این معیار مدلی که دارای کمترین مقدار  $DIC$  باشد به عنوان بهترین مدل انتخاب می‌شود. این معیار برای هر اندازه نمونه قابل استفاده بوده و با روش‌های زنجیر مارکوف مونت کارلویی به آسانی قابل محاسبه است (هستینگر، ۱۹۷۰).

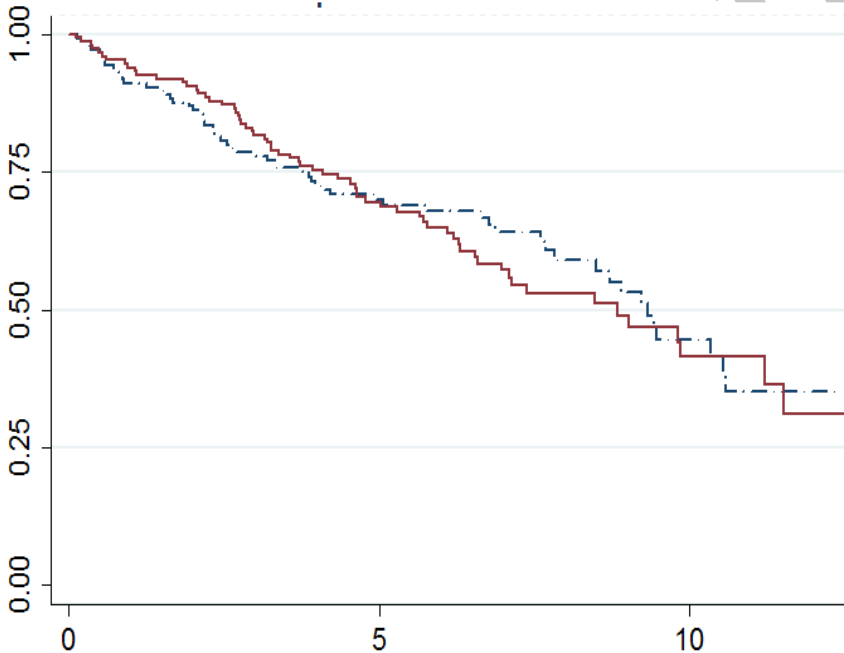
## ۵ مثال کاربردی

در مطالعه یک کارآزمایی بالینی دوسو کور ۴۲۴ بیمار مبتلا به سیروز کبدی طی سال‌های ۱۹۷۴ تا ۱۹۸۴ ثبت نام کردند. طبق پروتکل کارآزمایی و موافقت بیماران و پزشکان ۳۱۲ نفر به صورت تصادفی تحت درمان با دی پنسیلات (DPCA) یا دارو نما قرار گرفتند. برای این ۳۱۲ نفر تمام فاکتورهای کلینیکی، سربوژی و اتیولوژی جمع‌آوری گردید. تا زمان پایان پیگیری (۱۹۸۶) ۱۲۵ نفر از این ۳۱۲ نفر فوت کرده بودند که تنها ۱۱ نفر به دلیلی غیر از بیماری مورد نظر فوت کرده بودند و فقط پیگیری ۸ بیمار از دست رفته بود و ۱۹ نفر نیز تحت پیوند کبد قرار گرفتند که در این تحلیل حذف شده‌اند. از ۲۹۳ بیمار باقی مانده در این تحلیل ۱۴۸ نفر (۵۰/۵) تحت درمان با دارو قرار داشتند و ۸۸/۷۴ درصد از آنها مرد بودند. افراد شرکت کننده دارای حداقل ۲۷ سال و حداکثر ۷۹ سال با میانگین سنی ۵۰/۷۷ و انحراف استاندارد ۱۰/۶ بوده‌اند. کمترین زمان پیگیری برای بیماران در این مطالعه ۱/۵ ماه و بیشترین زمان پیگیری ۱۳ سال بوده است و ۱۱ درصد از بیماران پیگیری بیش از ۱۰ سال داشتند. ۵۷/۳۴ درصد از بیماران در این مطالعه سانسور شدند. در این کارآزمایی تعداد زیادی متغیر کمکی اندازه‌گیری شده بودند ولی در این تحلیل فقط متغیرهایی مورد بررسی قرار گرفته‌اند که از نظر بالینی مهم و به آسانی قابل اندازه‌گیری هستند. در نتیجه متغیرهای کمکی که در این تحلیل وارد شده‌اند عبارتند از درمان (=۰ دارونما، =۱ دارو) و سایر متغیرها مانند سن، بیلی روبین، آلومین و



پروترومبین به ترتیب با میانگین و انحراف معیار  $(۱۰/۶, ۵۰/۷۷)$ ،  $(۳/۲۶, ۴/۶۵)$ ،  $(۳/۵۲, ۰/۴۲)$  و  $(۱۰/۷۵, ۱/۰۲)$  (فلمینگ و هارینگتون، ۲۰۰۵).

یکی از روش‌های متداول و ساده برای شناسایی وجود افراد شفایافته رسم نمودار کاپلان-مایر است. در صورت وجود افراد شفایافته این نمودار قبل از رسیدن به نقطه صفر به صورت خط افقی در می‌آید (پراک و مانتوگا، ۲۰۰۱). همان‌طور که در شکل ۱ ملاحظه می‌شود در این داده‌ها وجود افراد مصون تایید



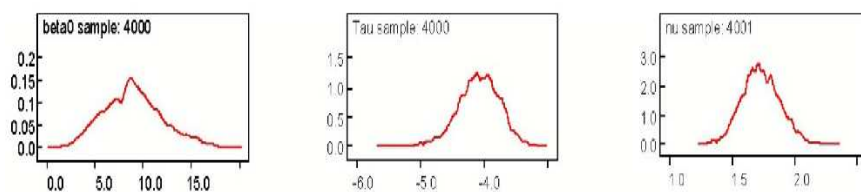
شکل ۱: نمودار کاپلان مایر متغیر درمان، دارو (خط پر) و دارونما (خط نقطه)

می‌شود. برای برآورد پارامترهای مدل، توزیع‌های پیشین

$$\tau \sim N(0, 1), \quad V \sim G(1, 1), \quad \lambda \sim G(1, 1), \quad p \sim G(1, 1)$$

برای پارامترهای مدل در نظر گرفته شده است. برای برازش مدل، برنامه مورد نظر در نرم‌افزار WinBUGS نوشته شده است (اشپیگل هالتر و همکاران، ۲۰۰۳). پس از تولید نمونه‌ها برای تشخیص همگرایی از آماره گلמן - روبین (بروک و گلמן، ۱۹۹۸) استفاده شده تا دوره سوخت مناسب تعیین گردد. از آنجایی که مقدار این

آماره برای همه پارامترها کمتر از ۱/۱ بود، تعداد ۱۰۰۰۰ نمونه برای دوره تطبیق مناسب به نظر می‌رسد. در نتیجه ۴۰۰۰۰ نمونه بعدی به عنوان نمونه‌های حاصل از توزیع پسین پارامترها در نظر گرفته شد و برای کاهش همبستگی از هر ۱۰ بار اجرا یک مرتبه نمونه‌گیری انجام گرفت. هموار بودن نمودارهای تابع چگالی پسین نشان دهنده همگرایی مناسب الگوریتم‌های زنجیر مارکوف مونت کارلو است. همچنین مشاهده می‌شود که نمودارهای تابع چگالی اثرات رگرسیونی تقریباً متقارن بوده اما نمودارهای تابع چگالی پارامتر  $\lambda$  دارای چولگی است. پایدار بودن نمودارهای اثر پارامترها در دامنه‌ای معین، و نبود افت و خیزهای طولانی در این نمودارها، نشان از همگرایی الگوریتم‌های عددی زنجیر مارکوف مونت کارلو است. نمودارهای تابع چگالی بعضی از پارامترها در شکل ۲ نشان داده شده است. خلاصه



شکل ۲: نمودار تابع چگالی برای پارامترهای زمان بقا پیشرفت (وایبول) و مقدار ثابت

آماره پسین پارامترهای مدل فوق هندسی تعمیم یافته دو جمله ای منفی، شامل برآوردهای نقطه‌ای (میانگین، انحراف معیار) و بازه باورمند ۹۵٪ پارامترها بر اساس نمونه با حجم ۴۰۰۰ تولید شده در جدول ۱ مشاهده می‌شود. از آنجایی که در مدل‌های شفایافته اثر متغیرهای کمکی بر روی نسبت شفایافتگی اندازه‌گیری می‌شود، تفسیر آنها با مدل‌های متداول بقا متفاوت است. علامت مثبت ضرایب رگرسیونی در این مدل شفایافته موجب افزایش نسبت شفایافتگی است، در صورتی که در مدل‌های متداول بقا علامت مثبت ضرایب رگرسیونی باعث افزایش خطر رخداد پیشامد می‌گردد. لذا با توجه به جدول ۱ ملاحظه می‌شود که متغیرهای میزان بیلی روبین و میزان آلبومین اثر منفی بر نسبت شفایافتگی دارند. به عبارت دیگر باعث کاهش نسبت شفایافتگی می‌شوند، متغیر پروترومبین دارای اثر مثبت بر نسبت

شفایافتگی است و اثر متغیر درمان نیز معنی دار نیست. همچنین متغیر سن نیز به‌طور جزئی معنی دار است و ضریب منفی آن بیانگر آن است که با افزایش سن نسبت شفایافتگی کاهش می‌یابد. نسبت شفایافتگی بر اساس متغیر درمان و سطوح

جدول ۱: برآورد پسینی میانگین، انحراف معیار و بازه باورمند ۹۵٪

پارامتر	میانگین	انحراف معیار	صدک ۲/۵	صدک ۹۷/۵
۳	-۴/۰۴۴	۰/۳۱۰	-۴/۶۷۸	-۳/۴۵۵
۷	۱/۷۲۱	۰/۱۴۷	۱/۴۴۹	۲/۰۲۲
مقدار ثابت	۸/۶۲۹	۳/۰۵۹	۲/۹۸۷	۱۴/۴۸۰
دارو	-۰/۲۱۸	۰/۳۳۰	-۰/۸۹۱	۰/۴۳۷
سن	-۰/۰۶۰	۰/۰۱۷	-۰/۰۹۶	-۰/۰۲۸
بیلی روبین	-۰/۵۲۷	۰/۱۵۲	-۰/۸۳۸	-۰/۲۴۵
پروترومبین	۱/۳۶۷	۰/۴۳۹	۰/۴۷۳	۲/۲۷۴
آلبومین	-۰/۷۴۲	۰/۲۵۹	-۱/۱۹۶	۰/۲۲۴
p	۰/۱۵۰	۰/۱۲۸	۰/۰۲۳	۰/۵۱۸
λ	۶/۳۱۳	۱/۸۵۸	۲/۹۳۳	۱۰/۳۵۰

مختلف بیلی روبین (چندک‌های ۲۵، ۵۰ و ۷۰) به‌شرط ثابت بودن سایر متغیرهای کمکی در مقدار میانگین‌شان در جدول ۴ نشان داده شده است. شایان ذکر است مقدار بیلی روبین برای چندک‌های ۲۵، ۲۵ و ۷۵ به ترتیب برابر ۰/۸، ۱/۳ و ۳/۴ است. همان‌طور که در جدول ۲ ملاحظه می‌شود با افزایش میزان بیلی روبین نسبت شفایافتگی در هر سه مدل کاهش می‌یابد. همچنین نسبت شفایافتگی در گروهی که دارو دریافت کرده‌اند بیشتر از گروه دارونما است، اگر چه این تفاوت معنی دار نیست. به‌نظر می‌رسد مدلی که نسبت شفایافتگی بالاتری را برآورد کرده است، بهتر به داده‌ها برازش داشته است. برای انجام مقایسه بین مدل ارائه شده در این مقاله با مدل‌های قبلی، مدل پواسون (مدل ۱) و مدل دو جمله‌ای منفی (مدل ۲) به داده‌ها برازش داده شد. همان‌طور که در جدول ۳ ملاحظه می‌شود مدل شفایافتگی با توزیع فوق هندسی تعمیم‌یافته دو جمله‌ای منفی (مدل ۳) با معیار اطلاع انحرافی ۷۴۷/۴۶ برازش بهتری در مقایسه با مدل پواسون و مدل دو جمله‌ای با معیار اطلاع انحرافی ۷۵۹/۵۸ و ۷۵۶/۳۸ به‌دست می‌دهد.

جدول ۲: برآورد نسبت شفایافتگی در سطوح مختلف متغیر بیللی روین برای مدل های مختلف

میزان بیللی روین						
۳/۴		۱/۳		۰/۸		مدل
دارو	دارونما	دارو	دارونما	دارو	دارونما	
۰/۶۳۱	۰/۵۵۶	۰/۷۲۰	۰/۶۸۳	۰/۸۱۲	۰/۷۸۵	۱
۰/۷۱۴	۰/۶۲۷	۰/۸۳۲	۰/۷۹۶	۰/۸۶۷	۰/۸۵۱	۲
۰/۸۰۶	۰/۷۳۸	۰/۹۶۵	۰/۹۵۰	۰/۹۷۷	۰/۹۶۷	۳

جدول ۳: معیار اطلاع انحرافی برای مدل های مختلف

مدل	DIC	PD	$\bar{D}(\theta)$
۱	۷۵۹/۵۸	۷/۷۹	۷۵۱/۷۹
۲	۷۵۶/۳۸	۷/۹۵	۷۴۸/۴۳
۳	۷۴۷/۴۶	۸/۰۴	۷۳۹/۴۲

### مطالعه شبیه سازی

برای بررسی دقت برآورد پارامترها، مطالعه ای شبیه سازی براساس مدل فوق هندسی تعمیم یافته دو جمله ای منفی با پارامتر زمان پیشرفت  $\nu = 2$  و  $\tau = 0.5$  (توزیع وایبول) انجام شده است. متغیر پنهان  $N$  از توزیع فوق هندسی تعمیم یافته دو جمله ای منفی با پارامترهای  $p$ ،  $\lambda$  و  $\alpha$  تولید شده است، به طوری که  $P = -1$ ،  $\lambda = 2$ ،  $\alpha = \exp(\beta_0 + \beta_1 X)$  و  $\beta_0 = -0.3$  و  $\beta_1 = 1$  در صورتی که متغیر  $X$ ، دو حالتی باشد و مقادیر آن از توزیع برنولی با  $\Pi = 0.5$  تولید شود. نسبت شفایافتگی در هر گروه برابر با  $38/5$  و  $29/5$  درصد به دست می آید. در صورتی که مقدار  $N$  صفر باشد، افراد شفایافته هستند در نتیجه زمان بقا برای این دسته از افراد طولانی خواهد بود و این مقدار زمان پایان مطالعه در نظر گرفته شده است. در صورتی که مقدار  $N$  مخالف صفر باشد به تعداد آن از توزیع وایبول با پارامترهای ذکر شده تولید می شود. به طوری که  $T = \min\{Y_1, \dots, Y_n\}$  با روش ارائه شده نمونه هایی به حجم ۲۵۰ و ۵۰۰ با تکرار ۱۰۰ از توزیع فوق هندسی تعمیم یافته

دوجمله‌ای منفی تولید می‌شود و برآورد پارامترهای مدل بر اساس سه مدل (فوق) هندسی تعمیم‌یافته دوجمله‌ای منفی و دوجمله‌ای منفی و پواسون) به دست آورده شده است. همان‌طور که در جدول ۴ ملاحظه می‌شود برآورد پارامترها در توزیع

جدول ۴: نتایج حاصل از شبیه‌سازی مدل  $P$ ،  $NB$  و  $HGNB$  به حجم نمونه  $250$  و  $500$  با تکرار  $100$

$n=500$		$n=250$		پارامتر	مدل
توان دوم	میانگین	توان دوم	میانگین		
خطا	خطا	خطا	خطا		
۰/۱۱۶	۰/۴۵	۰/۱۱۳	۰/۴۶	$\tau$	
۰/۱۷۱	۲/۱۵	۰/۱۶۳	۲/۱۳	$\nu$	
۰/۰۹۹	-۰/۴۱	۰/۰۹۵	-۰/۴۲	$\beta_0$	۱
۰/۱۳۱	۱/۱۶	۰/۱۲۱	۱/۱۵	$\beta_1$	
۰/۱۱۵	۰/۵۱	۰/۱۱۲	۰/۵۳	$\tau$	
۰/۰۹۶	۲/۵	۰/۰۹۱	۲/۱۱	$\nu$	
۰/۱۰۱	۰/۳۸	۰/۰۹۷	-۰/۴۱	$\beta_0$	۲
۰/۱۰۵	۱/۱	۰/۱۰۱	۱/۱	$\beta_1$	
۰/۱۳۶	۲/۱۰۸	۰/۴۵	۱/۰۵	$\lambda = p$	
۰/۱۰۵	۰/۵۱	۰/۱۰۷	۰/۵۱	$\tau$	
۰/۰۶۱	۲/۰۱	۰/۰۶۶	۲/۰۳	$\nu$	
۰/۰۶۲	-۰/۳۳	۰/۰۶۰	-۰/۳۷	$\beta_0$	۳
۰/۰۵۱	۰/۹۸	۰/۰۷۱	۰/۹۷	$\beta_1$	
۰/۲۶۱	۲/۰۲	۰/۲۸۰	۲/۰۲	$\lambda$	
۰/۰۷۰	-۱/۰۸	۰/۰۸۰	-۱/۱	$p$	

فوق هندسی تعمیم‌یافته دوجمله‌ای منفی نسبت به دو مدل دیگر به مقدار واقعی نزدیک‌ترند. با توجه به نتایج شبیه‌سازی روش ارائه شده برای تحلیل این گونه داده‌ها مورد تایید است و با افزایش حجم نمونه نتایج بهتری به دست می‌آید.

### بحث و نتیجه‌گیری

مدل جدیدی در رده مدل‌های شفایافته نامیخته تحت عنوان مدل‌های شفایافته زمان پیشرفت ارائه شد که شامل مدل دوجمله‌ای منفی، متداولترین توزیع در تبیین بیش

پراکنش داده‌ها، است. بر اساس معیار اطلاع انحرافی، میزان جریمه پیچیدگی برای مدل فوق هندسی تعمیم یافته دو جمله ای منفی بیشتر از مدل پواسون و دو جمله ای منفی است. بنابراین مدل دو جمله ای منفی در تبیین این بیش پراکنشی چندان موفق نبوده است. توزیع فوق هندسی تعمیم یافته دو جمله ای منفی وقتی  $\lambda = p$  باشد، به توزیع دو جمله ای منفی تقلیل می یابد. با توجه به برآوردهای این دو پارامتر فرض برابری آنها منتفی است و می تواند یکی از دلایل عدم برازش توزیع دو جمله ای منفی به این داده‌ها باشد. از آنجایی که این توزیع در حالت های خاص در بر گیرنده سایر توزیع ها است می توان از آن برای مقایسه بین مدل های مختلف استفاده کرد. مزیت دیگر مدل های شفایافته آن است که در صورت نبودن افراد شفایافته به مدل های متداول بقا تقلیل می یابند. مدل بقای متداول (بدون نسبت شفایافتگی) با تابع بقا وایبول برازانده شده به داده‌ها در مقایسه با مدل های شفایافته نشان دهنده برازش ضعیف مدل متداول بقا به این داده‌ها ( $DIC=۸۶۱/۹۱$ ) است.

## منابع

- Baghestani, A. R., Rahimzadeh, M. and Pourhosseingoli, M, A. (2013), Generalized Poisson- Lindely Distribution in Promotion Time Cure Model, *Pakistan Journal of Statistics* (Submitted).
- Boag, J. W. (1949), Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy, *Journal of the Royal Statistics Society*, B, **11**, 15-44.
- Borges, P., Rodrigues, J. and Balakrishnan, N. (2012), Correlated Destructive Generalized Power Series Cure Rate Models and Associated Inference with an Application to a Cutaneous Melanoma Data, *Computational Statistics Data Analysis*, **56**, 1703-1713.

- Brooks, S. and Gelman, A. (1998), General Methode for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphics Statistics*, **8**, 319-335.
- Cancho, V. G., Rodrigues, J. and De Castro, M. (2011), A Flexible Model for Survival Data with a Cure Rate: A Bayesian Approach, *Journal of Applied Statistics*, **38**, 57-70.
- Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999), A New Bayesian Model for Survival Data with a Surviving Fraction. *Journal of the American Statistical Association*, **94**, 909-919.
- Cooner, F., Banerjee, S., Carlin, B. P. and Sinha, D. (2007), Flexible Cure Rate Modeling Under Latent Activation Schemes, *Journal of the American Statistical Association*, **102**, 560-572.
- Fleming, T. R., Harrington, D. P. (2005), Counting Processes and Survival Analysis, John Wiley & Sons.
- Ghitani, M. E., AL-Awadhi S. A. and Kalla S. L. (2002), On Hypergeometric Generalized Negative Binomial Distribution. *The International Journal of Mathematics and Mathematical Sciences*, **29**, 727-736.
- Hastings W. K, (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, **57**, 97-109.
- Maller, R. A. and Zhou, S. (1996), *Survival Analysis with Long-Term Survivors*, John Wiley & Sons, U.K.
- Price, D. L. and Manatunga, A. K. (2001), Modeling Survival Data with a Cured Fraction Using Frailty Models, *Statistics in Medicine*, **20**, 1515-1527.

Rodrigues J., Castrode M., Cancho V. G. and Balakrishnan N. (2009), COM-Poisson Cure Rate Survival Models and an Application to a Cutaneous Melanoma Data, *Journal of Statistical Planning and Inference*, **139**, 3605-3611.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002), A Bayesian Measures of Model Complexity and Fit, *Journal of Royal Statistical Society, B*, **64**, 583-639.

Spiegelhalter, D. J., Thomas, A., Best, N. and Lunn, D. (2003), WinBUGS User Manual, Version 1.4, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College, UK., available at: <http://www.mrcbsu.cam.ac.uk/bugs>.

Yakovlev, A. Y. and Tsodikov, A. D. (1996), Stochastic Models of Tumor Latency and Their Biostatistical Applications, World Scientific, Singapore, Hackensack.

Archive of SID