انتخاب مدل وایبول یا مدل گاوسی وارون بر اساس آزمون نسبت درستنمایی

رضا پاکیاری[۱]

چکیده

در این مقاله به کمک آزمون نسبت درستنمایی، روشی برای انتخاب یکی از دو مدل وایبول و یا گاوسی وارون ارائه شده است. به کمک شبیه سازی مونت کارلو، احتمال انتخاب صحیح و احتمالهای خطاهای نوع اول و دوم در این روش محاسبه شده اند و دیدیم که این آزمون برای نمونه های به اندازهٔ ۱۰۰ و بیشتر، از کارایی مناسبی برخوردار است و سرانجام این روش را برای چند مجموعه از داده های موجود به کار برده ایم.

واژه های کلیدی: توزیع وایبول، توزیع گاوسی وارون، آزمون نسبت درستنمایی، احتمال انتخاب صحیح.

۱ پیشگفتار

مدلهای وایبول و گاوسی وارون از مدلهای مهم آماری در مبحث قابلیت اعتماد و مسائل مربوط به آزمون های طول عمرند. انتخاب صحیح یکی از این دو مدل برای توصیف داده های جمع آوری شده در اعتبار نتایج به دست آمده از استنباط آماری از اهمیت بسیاری برخوردار است.

توزیع وایبول به طوری گسترده در توصیف داده های مربوط به طول عمر به کارگرفته شده است و اشخاص زیادی خصوصیات این توزیع را بررسی کرده اند. بحث کاملی از این توزیع را می توان در [۳] یافت. توزیع گاوسی وارون نیز دارای کاربردهای وسیعی در قابلیت اعتماد است و بحث مفصلی از این توزیع را می توان در [۲] مطالعه کرد. ما در این مقاله به کمک آزمون نسبت درستنمایی روشی را برای انتخاب یکی از این دو مدل ارائه می کنیم. بین و انگلهارت[۲] [۱] از این روش برای مسئله انتخاب مدل وایبول یا مدل گاما استفاده کرده اند. در بخش ۲ به معرفی توزیع وایبول و برآوردگرهای درستنمایی ماکسیمم پارامترهای آن و هم چنین تولید کردن عدد تصادفی از این توزیع پرداخته ایم و در بخش ۳ به معرفی توزیع گاوسی وارون می پردازیم. آزمون نسبت درستنمایی روش مناسبی را برای انتخاب یکی از این دو توزیع بر اساس داده های موجود به ما می دهد و ما در بخش ٤ این آزمون و قاعدهٔ تصمیم گیری آن را ارائه خواهیم کرد. در بخش ٥ اعتبار آزمون به کمک شبیه سازی مونت کارلو و به طریق محاسبهٔ احتمال انتخاب صحیح بررسی شده است و بر اساس نتایج به دست آمده معلوم می شود که حتی برای نمونه های کوچک، آزمون از اعتبار نسبتاً مناسبی برخوردار است و سرانجام در بخش ٦ روش ارائه شده را برای دو مثال متداول به کار برده ایم.

---

[۱] گروه ریاضی ـ دانشگاه اراک

[۲] Bain and Engelhardt

۲ توزیع وایبول

تابع چگالی توزیع وایبول دو پارامتری به صورت زیر است:

$$f_1(x;b,c)=\frac{c}{b}\left(\frac{x}{b}\right)^{c-1}\exp\left[-\left(\frac{x}{b}\right)^{c}\right] \qquad (۱)$$

$$;c>0,b>0,x>0$$

که در آن $b>0$ پارامتر مقیاس و $c>0$ پارامتر شکل توزیع است.

برآوردگرهای درستنمایی ماکسیمم پارامترهای b و c توسط روابط زیر داده می شوند: [۸]

$$\frac{1}{\hat{c}}-\frac{\sum_{i=1}^{n}x_i^{\hat{c}}\ln x_i}{\sum_{i=1}^{n}x_i^{\hat{c}}}+\frac{\sum_{i=1}^{n}\ln x_i}{n}=0 \qquad (۲)$$

و

$$\hat{b}=\left(\frac{\sum_{i=1}^{n}x_i^{\hat{c}}}{n}\right)^{1/\hat{c}} \qquad (۳)$$

ابتدا از معادله (۲)، مقدار $\hat{c}$ با یک روش تکراری مانند روش نیوتن ـ رافسون محاسبه می شود و سپس از معادله (۳) مقدار $\hat{b}$ به دست می آید. همچنین تابع توزیع وایبول دو پارامتری به صورت زیر است:

$$F(x)=1-\exp\left[-\left(\frac{x}{b}\right)^{c}\right] \qquad ; x>0 \qquad (٤)$$

با معکوس کردن تابع توزیع فوق و به کمک قضیه تبدیل انتگرال احتمال [۷؛ صفحهٔ ۲۰۲] نتیجه می شود که هرگاه RND عدد تصادفی از بازهٔ $(0,1)$ باشد آنگاه: $b[-\ln(1-RND)]^{1/c}$ مقداری از توزیع وایبول با پارامترهای (b,c) خواهد بود. ما از این روش برای تولید مشاهدات تصادفی از توزیع وایبول به منظور شبیه سازی این توزیع استفاده کرده ایم.

۳ توزیع گاوسی وارون

تابع چگالی گاوسی وارون به صورت زیر است:

$$f_2(x;\mu,\lambda)=\sqrt{\frac{\lambda}{2\pi}}\,x^{-3/2}\exp\left[\frac{-\lambda(x-\mu)^2}{2\mu^2x}\right] \qquad (۵)$$

$$;\lambda>0,\mu>0,x>0$$

برآوردگرهای درستنمایی ماکسیمم پارامترهای μ و λ توسط روابط زیر داده می شوند: [۲]

$$\hat{\mu}=\overline{X} \qquad (٦)$$

$$\frac{1}{\hat{\lambda}}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{X_i}-\frac{1}{\overline{X}}\right) \qquad (۷)$$

متأسفانه معکوس تابع توزیع گاوسی وارون به سادگی به دست نمی آید و بنابراین از قضیه تبدیل انتگرال احتمال نمی توان عدد تصادفی از این توزیع تولید کرد.

مایکل و همکارانش [٦] روشی را برای تولید توزیع گاوسی وارون به کمک تبدیلی با دو ریشه به صورت زیر ارائه کرده اند:

فرض کنید $X\sim IG(\mu,\lambda)$ ثابت می شود متغیر تبدیل یافته $Y^2=\frac{\lambda(X-\mu)^2}{\mu^2X}$ دارای توزیع $\chi^2_{(1)}$ با دو ریشهٔ X_1 و X_2 به صورت زیر است:

$$X_1=\frac{\mu}{2\lambda}\left[2\lambda+\mu Y^2-\sqrt{4\lambda\mu Y^2+\mu^2Y^4}\right] \qquad (۸)$$

$$X_2=\frac{\mu^2}{X_1} \qquad (۹)$$

برای یک مقدار Y^2 ی تولید شده از توزیع $\chi^2_{(1)}$، مایکل و همکارانش احتمال شرطی انتخاب هر ریشه را حساب کردند. ریشهٔ کوچکتر X_1 با احتمال $\frac{\mu}{\mu+x_1}$ و ریشهٔ دیگر X_2 با احتمال $\frac{x_1}{\mu+x_1}$ انتخاب می شود. بنابراین دستور العملی برای تولید یک مشاهده تصادفی از توزیع گاوسی وارون به صورت زیر است:

۱ـ یک عدد تصادفی از توزیع مربع خی با یک درجهٔ آزادی تولید کنید.

۲ـ برای عدد تصادفی به دست آمده در مرحلهٔ ۱، ریشهٔ کوچکتر یعنی x_1 را به کمک رابطهٔ (۸) محاسبه کنید.

۳ـ یک آزمایش برنولی با احتمال موفقیت $p=\frac{\mu}{\mu+x_1}$ را انجام دهید.

۴ـ هرگاه نتیجهٔ آزمایش برنولی موفقیت باشد x_1 به عنوان یک مشاهدهٔ تصادفی از توزیع $IG(\mu,\lambda)$ انتخاب می شود و در غیر این صورت ریشهٔ دیگر یعنی x_2 انتخاب می شود.

۴ آزمون نسبت درستنمایی

فرض کنید $x_1,x_2,\dots,x_n$، یک نمونهٔ تصادفی از توزیع $F(x,\theta)$ باشد. علاقه مند به انجام آزمون فرض زیر هستیم:

$$\begin{cases} H_0: X_i \sim W(b,c) \\ H_1: X_i \sim IG(\mu,\lambda) \end{cases} \qquad (10)$$

به کمک آزمون نسبت درستنمایی، ناحیهٔ بحرانی را برای فرض فوق به دست خواهیم آورد.

فرض کنید $L^*_\omega=\ln L_\omega(\hat b,c)$ که در آن $L_\omega(\hat b,c)$ تابع درستنمایی توزیع وایبول به ازای c و $\hat b$ است. همچنین فرض کنید $L^*_I=\ln L_I(\hat\mu,\hat\lambda)$ که در آن $\ln L_I(\hat\mu,\hat\lambda)$ تابع درستنمایی توزیع گاوسی وارون به ازای $\hat\mu$ و $\hat\lambda$ است. در این صورت با توجه به توابع چگالی (۱) و (۵) داریم:

$$L^*_\omega = n\ln\hat c - n\hat c\ln\hat b + (\hat c-1)\sum_{i=1}^{n}\ln x_i - \sum_{i=1}^{n}\left(\frac{x_i}{\hat b}\right)^{\hat c} \qquad (11)$$

$$L^*_I = \frac{n}{2}\ln\hat\lambda - \frac{n}{2}\ln(2\pi) - \frac{3}{2}\sum_{i=1}^{n}\ln x_i - \lambda\sum_{i=1}^{n}\left[\frac{(x_i-\bar x)^2}{2\bar x^2 x_i}\right] \qquad (12)$$

بنابراین هرگاه آمارهٔ T را لگاریتم طبیعی نسبت درستنمایی تعریف کنیم یعنی هرگاه

$$T=\ln\left[\frac{L^*_I(\hat\mu,\hat\lambda)}{L^*_\omega(\hat b,\hat c)}\right] \qquad (13)$$

خواهیم داشت:

$$T=\frac{n}{2}\ln\hat\lambda - \frac{n}{2}\ln(2\pi) - (\hat c+\frac{1}{2})\sum_{i=1}^{n}\ln x_i - \hat\lambda\sum_{i=1}^{n}\left[\frac{(x_i-\bar x)^2}{2\bar x^2 x_i}\right] - n\ln\hat c + n\hat c\ln\hat b + \sum_{i=1}^{n}\left(\frac{x_i}{\hat b}\right)^{\hat c}$$

پس از کمی ساده سازی نتیجه می شود که:

$$T=\frac{n}{2}[\ln\hat\lambda - \ln(2\pi) - 2\ln\hat c + 2\hat c\ln\hat b] - (\hat c+\frac{1}{2})\times \sum_{i=1}^{n}\ln x_i + \sum_{i=1}^{n}\left(\frac{x_i}{\hat b}\right)^{\hat c} - \hat\lambda\left[\frac{n}{2\bar x} - \frac{n}{\bar x} + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)\right] \qquad (14)$$

بنابراین برای داده های مشاهده شدهٔ $x_1,x_2,\dots,x_n$ نخست به کمک روابط (۲) و (۳) و (٦) و (۷)، مقادیر $\hat\lambda,\hat\mu,\hat b,\hat c$ را محاسبه کرده و آنگاه T را از رابطهٔ فوق به دست می آوریم. سپس از قاعدهٔ تصمیم زیر استفاده می کنیم:

اگر $T>0$ باشد مدل گاوسی وارون را به عنوان مدل مناسب برای داده ها انتخاب کنید، در غیر این صورت مدل وایبول را انتخاب کنید.

۵ شبیه سازی

برای بررسی اعتبار آزمون به روش شبیه سازی مونت کارلو نمونه هایی را از توزیع وایبول و توزیع گاوسی وارون تولید کرده ایم و احتمال انتخاب صحیح را برای هر توزیع به کمک ۲۰۰۰ نمونه محاسبه نموده ایم. نتایج بدست آمده در جداول زیر درج شده اند. توجه کنید که هرگاه نمونهٔ تصادفی از توزیع وایبول انتخاب شده باشد، احتمال انتخاب صحیح یعنی $P[T<0\mid x_i\sim W(b,c)]$ فقط به پارامتر c بستگی دارد، زیرا با توجه به رابطهٔ (۳)، MLE پارامتر b بستگی به پارامتر c دارد و در نتیجه تابع درستنمایی تنها به پارامتر c بستگی دارد. همچنین هرگاه نمونهٔ تصادفی از توزیع گاوسی وارون استخراج شده باشد، احتمال انتخاب صحیح یعنی $P[T>0\mid x_i\sim IG(\mu,\lambda)]$ فقط به پارامتر λ بستگی خواهد داشت، زیرا همان طور که در رابطهٔ (۱۲) مشاهده می شود تابع درستنمایی به پارامتر μ بستگی ندارد.

بر پایهٔ احتمال انتخاب صحیح، می توان احتمالهای خطاهای نوع اول و نوع دوم را نیز به دست آورد. توجه کنید که:

$$\alpha = P(\text{Reject } H_0 \mid H_0 \text{ true}) = P[T > 0 \mid X_i \sim W] = 1 - P[T \leq 0 \mid X_i \sim W]$$

$$\beta = P(\text{Reject } H_1 \mid H_1 \text{ true}) = P[T < 0 \mid X_i \sim IG] = 1 - P[T \geq 0 \mid X_i \sim IG] \quad (15)$$

بنابراین متمم احتمالهای ثبت شده در جدول ۱ و ۲ به ترتیب احتمالهای خطاهای نوع اول و نوع دوم را به دست می دهند که در جداول ۳ و ٤ درج شده اند.

همانطور که از جداول ۱ تا ٤ مشاهده می شود با افزایش نمونه، احتمال انتخاب صحیح افزایش یافته و به طور متناظر احتمالهای خطاهای نوع اول و نوع دوم کاهش می یابند. به طور مشخص برای $n = ۱۰۰$ احتمالهای خطاهای نوع اول و نوع دوم کوچک بوده و آزمون از کارایی مناسبی برخوردار است.

٦ مثالها

این روش را برای دو مجموعه از داده ها به کار برده ایم و نتایج جالبی را به دست آورده ایم. لیبلین[3] و زلن[4] [٥]، $n = ۲۳$ داده زیر را برای توصیف مدل وایبول به کار برده اند و پس از آنان مؤلفان بسیاری از این داده ها به عنوان مثالی از توزیع وایبول استفاده کرده اند (به عنوان مثال [٦] را ببینید).

۱۷/۸۸	۲۸/۹۲	۳۳/۰۰	٤۱/۵۲	٤۲/۱۲
٤۵/٦۰	٤۸/٤۸	۵۱/۸٤	۵۱/۹٦	۵٤/۱۲
۵۵/۵٦	٦۷/۸۰	٦۸/٦٤	٦۸/٦٤	٦۸/۸۸
۸٤/۱۲	۹۳/۱۲	۹۸/٦٤	۱۰۵/۱۲	۱۰۵/۸٤
۱۲۷/۹۲	۱۲۸/۰٤	۱۷۳/٤۰		

برای داده های فوق داریم:

$$\hat{\lambda} = ۲۳۱/۷۲ \text{ ، } \hat{\mu} = ۷۲/۲۲ \text{ ، } \hat{b} = ۸۱/۹۹ \text{ ، } \hat{c} = ۲/۱۰۲$$

و آمارهٔ نسبت درستنمایی به صورت $T = ۰/٤۹ > ۰$ به دست می آید. بنابراین باید مدل گاوسی وارون را به جای مدل وایبول برای داده های فوق به کار برد. البته در این مثال n نسبتاً کوچک است و از جدول ۳ در می یابیم که احتمال ارتکاب خطای نوع اول برابر ۰/۳۷ است.

گراس[5] و کلارک[6] [٤] داده های زیر را در نظر گرفته اند که مربوط است به مدت زمان تسکین ۲۰ بیمار که یک داروی مسکن را مصرف کرده اند. آنها از این داده ها برای تشریح روشهای توزیع وایبول سود جسته اند:

۱/٦	۱/۱	۱/٤	۱/۳	۱/۷	۱/۹	۱/۸	۱/٦	۲/۲	۱/۷
۲/۰	۲/۷	٤/۱	۱/۸	۱/۵	۱/۲	۱/٤	۳/۰	۱/۷	۲/۳

برای این داده ها داریم:

$$\hat{\lambda} = ۱۸/۷۰ \text{ ، } \hat{\mu} = ۱/۹۰ \text{ ، } \hat{b} = ۲/۱٤ \text{ ، } \hat{c} = ۲/۷۹$$

و لذا $T = ۳/۸ > ۰$. مشاهده می شود که در این حالت نیز باید از توزیع گاوسی وارون به جای توزیع وایبول استفاده کرد. در اینجا نیز n خیلی بزرگ نیست و از جدول ۳ مقدار تقریبی احتمال ارتکاب خطای نوع اول مجدداً ۰/۳۷ به دست می آید.

بین و انگلهارت [۱] نیز این دو مثال را مورد بررسی قرار داده اند و در هر دو مورد به این نتیجه رسیده اند که داده ها را نمی توان از توزیع وایبول تلقی کرد.

در پایان از گروه پژوهشی دانشگاه اراک به خاطر حمایت مالی از این اثر و همچنین از داوران محترم به خاطر نظرات سودمندشان تشکر می شود.

---

[3] Lieblein

[4] Zelen

[5] Gross

[6] Clark

جدول ۱

احتمال انتخاب صحیح برای مدل وایبول ; P[T <۰]

n \ c	۰/۵	۱	۲	۵	۱۰
۱۰	۰/۵۷	۰/۵۵	۰/٦۳	۰/٦۵	۰/٦۷
۲۰	۰/۵۹	۰/۵٦	۰/٦۳	۰/٦۸	۰/٦۸
۵۰	۰/۷۵	۰/٦۵	۰/۷۳	۰/۸۱	۰/۸۵
۱۰۰	۰/۹۱	۰/۹۰	۰/۹۳	۰/۹۸	۰/۹۹

جدول ۲

احتمال انتخاب صحیح برای مدل گاوسی وارون ; P[T >۰]

n \ λ	۱	۵	۱۰	۲۰	۵۰
۱۰	۰/٦۰	۰/۵٤	۰/۵۹	۰/٦٦	۰/٦۵
۲۰	۰/٦۵	۰/۵۹	۰/٦٤	۰/۷۱	۰/۷۲
۵۰	۰/۷۷	۰/٦٦	۰/۷۸	۰/۸۵	۰/۸۵
۱۰۰	۰/۹۲	۰/۸۹	۰/۹۵	۰/۹۷	۰/۹۸

جدول ۳

احتمال خطای نوع اول

n \ c	۰/۵	۱	۲	۵	۱۰
۱۰	۰/٤۳	۰/٤۵	۰/۳۷	۰/۳۵	۰/۳۳
۲۰	۰/٤۱	۰/٤٤	۰/۳۷	۰/۳۲	۰/۳۲
۵۰	۰/۲۵	۰/۳۵	۰/۲۷	۰/۱۹	۰/۱۵
۱۰۰	۰/۰۹	۰/۱۰	۰/۰۷	۰/۰۲	۰/۰۱

جدول ٤

احتمال خطای نوع دوم

n \ λ	۱	۵	۱۰	۲۰	۵۰
۱۰	۰/٤۰	۰/٤٦	۰/٤۱	۰/۳٤	۰/۳۵
۲۰	۰/۳۵	۰/٤۱	۰/۳٦	۰/۲۹	۰/۲۸
۵۰	۰/۲۳	۰/۳٤	۰/۲۲	۰/۱۵	۰/۱۵
۱۰۰	۰/۰۸	۰/۱۱	۰/۰۵	۰/۰۳	۰/۰۲

مراجع

[1] Bain , L.J. , Engelhardt, M.(1980). *Probability of Correct Sellection of Weibull Versus Gamma Based on Likelihood Ratio.commun.stat.theory and methods.*, 375-381.

[2] Chhikara, R.S. , Folks, J.L.(1989). *The Inverse Gaússian Distribution*, Marcel Dekker, Inc.

[3] Cohen, A.C. , Whitten, B.J.(1988). *Parameter Estimation in Reliability and Life Span Models*. Marcel Dekker, Inc.

[4] Gross, A.J. , Clark, V.A.(1975). *Survival Distribution: Reliability Applicationin in the Biomedical Science* . John Wiley and sons, Inc.

[5] Lieblein , J. Zelen , M.(1956). *Statistical Investigation of the Fatigue Life of Deepgroove Ball Bearings*, J.Res.Nat.Bar.Stand. 57, 273-316.

[6] Micheal, J.R. , Schucany , W.R. , Hass , R.W.(1976). *Generating Random Variables Using Transformation with Multiple Roots*, Amer.Statist. 30; 88-90.

[7] Mood,A.M., Graybill,F.A., Boes,D.C.(1974). *Introduetion to the Theory of Statistics.*, Third Edition . Mcgrawhill .

[8] Thoman,D.R., Bain,L.J., Antie,C.E.(1969). *Inferences on the Parameters of the Weibull Distribution.* Technometrics 11, 445-460.