

اندیشه آماری، بهار و تابستان ۱۳۹۱، شماره پیاپی ۳۳

سال هفدهم شماره اول، ص ۷۰-۸۶

توزیع t -چوله جایگزین مناسبی برای توزیع نمایی وزنی

محمد مهدی مقامی^۱، محمد بهرامی^۲

چکیده:

در این مقاله، ابتدا به معرفی اجمالی توزیع‌های t -چوله و نمایی وزنی پرداخته و برخی از خواص مهم آن‌ها را بررسی می‌کنیم. سپس با استفاده از مجموعه داده‌های واقعی که برای نشان دادن برتری توزیع نمایی وزنی بر توزیع‌های وایبل، گاما و نمایی تعمیم یافته استفاده شده است، نشان می‌دهیم که داده‌ها توسط توزیع t -چوله نسبت به توزیع نمایی وزنی بهتر برازش می‌شوند. سرانجام ادعای خود را با استفاده از شبیه سازی نیز بررسی می‌کنیم.

واژه‌های کلیدی: توزیع t -چوله، توزیع نمایی وزنی، p -مقدار، لگاریتم درستنمایی.

^۱گروه آمار دانشگاه اصفهان

^۲استادیار گروه آمار دانشگاه اصفهان

۱ مقدمه

۲ توزیع t -چوله

اولین بار آزالینی (۱۹۸۵) توزیع نرمال چوله را با اضافه کردن پارامتر چولگی به توزیع نرمال مطرح کرد. توزیع t -چوله^۳ حالت کلی تری از توزیع نرمال چوله است و به مدل اجازه می‌دهد، علاوه بر چولگی دارای دم نیز باشد. ساخت توزیع t از طریق

$$Z_{\lambda} \sim SN(\lambda)$$

نسبت یک توزیع نرمال مستقل به تبدیلی از متغیر خی - دو است. برای ساخت توزیع t -چوله همین ایده به کار می‌رود (آزالینی و کاپیتانیو (۲۰۰۳)). از طرفی گوپتا و کاندو (۲۰۰۹) در رده جدیدی از

توزیع‌های نمایی وزنی^۴ پارامتر چولگی را به توزیع

$$X | Y < \lambda X \sim SN(\lambda)$$

نمایی اضافه کردند. سپس به بررسی ارتباط این

$$\begin{aligned} f_{X|Y<\lambda X}(z, \lambda) \\ = \frac{f_X(z)P(Y < \lambda X | X = z)}{P(Y < \lambda X)} \end{aligned}$$

رده با توزیع‌های نمایی، گاما و نمایی تعمیم یافته (گوپتا و کاندو (۲۰۰۷)) پرداخته و با مثال‌هایی از داده‌های واقعی نشان دادند که رده مطرح شده نسبت به توزیع‌های وایبل، گاما و نمایی تعمیم یافته دارای برتری است. بخش‌های ۲ و ۳ به ترتیب به معرفی توزیع‌های t -چوله و نمایی وزنی می‌پردازند.

در بخش (۱.۴) با استفاده از مثال‌های واقعی گوپتا و کاندو (۲۰۰۹) و معیارهای متفاوت مقایسه دو

مدل، نشان می‌دهیم که مدل t -چوله به داده‌ها برازنده تر است. شبیه‌سازی‌های انجام شده در بخش (۴.۲)، نتایج بخش (۱.۴) را تایید می‌کند.

تعریف ۳.۲. گویم متغیر تصادفی $W_{\nu, \lambda}$ دارای

دارای توزیع t -چوله با پارامترهای $\nu > 0$ و $\lambda \in R$

است اگر $Z_{\lambda} = \frac{Z_{\lambda}}{\sqrt{\nu}}$ که در آن $W_{\nu, \lambda} \sim$

$SN(\lambda)$ و $V = \frac{\chi_{\nu}^2}{\nu}$ مستقل از Z_{λ} است. این

توزیع را با نماد $W_{\nu, \lambda} \sim St(\nu, \lambda)$ نشان می‌دهیم.

^۳Skew t distribution

^۴Weighted exponential distribution

تعریف ۴.۲. متغیر تصادفی $W = (W_1, \dots, W_n)^T$

دارای توزیع n متغیره t -استیودنت با پارامترهای

μ, Σ و درجه آزادی ν است، اگر تابع چگالی آن

به صورت

$$f(x | \mu, \Sigma, \nu) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{n}{2}}} |\Sigma|^{-\frac{1}{2}}$$

$$\times \left\{ 1 + \frac{(x - \mu)' \Sigma^{-1} (x - \mu)}{\nu} \right\}^{-(\frac{\nu+n}{2})}$$

باشد که $x \in R^2$.

این توزیع را با نماد $t_k(\mu, \Sigma, \nu)$ تابع چگالی آن

را با نماد $t_k(x; \mu, \Sigma, \nu)$ و تابع توزیع آن را با

نماد $T_k(x; \mu, \Sigma, \nu)$ نشان می‌دهیم.

برای تولید اعداد تصادفی از توزیع t -چوله

می‌توان از لم زیر استفاده کرد.

لم ۵.۲. اگر

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$n - m$ باشد و μ و Σ به صورت

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

افراز شده باشند، آنگاه:

و

$$(W_1, W_2)^T \sim t_2(0, I_2, \nu)$$

آنگاه

$$W_1 | W_2 < \lambda W_1 \sim St(\nu, \lambda).$$

الف-

$$W_1 \sim t_m(\mu_1, \Sigma_{11})$$

$$W_2 \sim t_m(\mu_2, \Sigma_{22})$$

اثبات. می‌دانیم $\frac{(X, Y)}{\sqrt{V}} \stackrel{d}{=} (W_1, W_2)$ که در

آن $(X, Y) \sim N_2(0, I_2)$ و $V \sim \chi_\nu^2/\nu$.

بنابراین با توجه به لم ۲.۲ می‌توان نوشت

$$\begin{aligned} & W_1 | W_2 < \lambda W_1 \\ \stackrel{d}{=} & \frac{X}{\sqrt{V}} | \frac{Y}{\sqrt{V}} < \lambda \frac{X}{\sqrt{V}} \\ \stackrel{d}{=} & \frac{X}{\sqrt{V}} | Y \\ \stackrel{d}{=} & \frac{X | Y < \lambda X}{\sqrt{V}} \\ \stackrel{d}{=} & \frac{Z_\lambda}{\sqrt{V}} \\ \stackrel{d}{=} & St(\nu, \lambda) \end{aligned}$$

□

پس اثبات کامل است.

قضیه ۶.۲. اگر

$$W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \sim t_n(\mu, \Sigma, \nu)$$

به طوری که بعد W_1 و W_1 به ترتیب برابر m و

ب- از طرفی با استفاده از قسمت (الف) قضیه ۶.۲،

$$W_1 \sim t(\nu) \text{ بنابراین}$$

$$W_1 | W_2 = w_2 \sim$$

$$\begin{aligned} & f_{St}(y; \nu, \lambda) \\ &= \frac{t(y; \nu) P[W_2 < \lambda W_1 | W_1 = y]}{P[W_2 < \lambda W_1]} \end{aligned} \quad \begin{aligned} & t_m(\mu_1(w_2), \Sigma_{11}(w_2), \nu + n - m) \\ & \mu_1(w_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (w_2 - \mu_2) \end{aligned}$$

اکنون از نتیجه ۷.۲ داریم: و $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ که در آن

$$W_2 | W_1 = y \sim t(0, \frac{\nu + y^2}{\nu + 1}) \quad \Sigma_{11}(w_2) = \frac{\nu + (w_2 - \mu_2)^T w_2 - \mu_2}{\nu + n - m}$$

یا نتیجه ۷.۲ در حالت خاص اگر

$$(\sqrt{\frac{\nu+1}{\nu+y^2}} W_2) | W_1 = y \sim t(0, 1, \nu+1) \quad (W_1, W_2)^T \sim t_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, I_{2 \times 2}, \nu \right)$$

پس

چون $n = 2$ و $m = 1$ ، آنگاه داریم:

$$\begin{aligned} & f_{St}(y; \nu, \lambda) = t(y; \nu) \times \\ & \frac{P[\sqrt{\frac{\nu+1}{\nu+y^2}} W_2 < \lambda \sqrt{\frac{\nu+1}{\nu+y^2}} | W_1 = y]}{P[W_2 < \lambda W_1]} \end{aligned} \quad W_1 | W_2 = w_2 \sim t_1(0, \frac{\nu + w_2^2}{\nu + 1}, \nu + 1)$$

یا

$$(\sqrt{\frac{\nu+1}{\nu+w_2^2}} W_1) | W_2 = w_2 \sim t_1(0, 1, \nu+1)$$

سرانجام از تقارن W_1 و W_2 حول صفر داریم:

$$P[W_2 < \lambda W_1] = \frac{1}{2} \text{ بنابراین}$$

قضیه ۸.۲. تابع چگالی توزیع t -چوله به صورت

$$\begin{aligned} & f_{St}(y; \nu, \lambda) \\ &= 2t(y; \nu) T[\sqrt{\frac{\nu+1}{\nu+y^2}} \lambda y; \nu + 1] \end{aligned} \quad f_{St}(y; \nu, \lambda) = 2t(y; \nu) T[\sqrt{\frac{\nu+1}{\nu+y^2}} \lambda y; \nu + 1]$$

به دست می آید.

و اثبات تمام است. $\square$

اثبات. فرض کنید $(W_1, W_2)^T \sim t_2(0, I_2, \nu)$.

تعریف ۹.۲. فرض کنید $W_{\nu, \lambda} \sim St(\nu, \lambda)$.

با استفاده از لم ۵.۲ می توان نوشت:

در این صورت $Y = \mu + \sigma W_{\nu, \lambda}$ دارای توزیع t -چوله با پارامترهای $(\mu, \sigma, \lambda, \nu)$ بوده و آن را با نماد $Y \sim St(\mu, \sigma, \lambda, \nu)$ نشان می دهیم. با در نظر گرفتن $\frac{Y - \mu}{\sigma} \sim St(\lambda, \nu)$ تابع چگالی Y به

$$\begin{aligned} & f_{St}(y; \nu, \lambda) \\ &= f_{W_1 | W_2 < \lambda W_1}(y; \nu, \lambda) \\ &= \frac{f_{W_1}(y) P(W_2 < \lambda W_1 | W_1 = y)}{P[W_2 < \lambda W_1]} \end{aligned}$$

صورت زیر است:

و استقلال این دو داریم:

$$f(y; \mu, \sigma, \lambda, \nu) = 2t\left(\frac{y-\mu}{\sigma}; \nu\right) T\left\{\lambda \frac{y-\mu}{\sigma} \left(\frac{\nu+1}{Q_y+\nu}\right); \nu+1\right\}$$

که در آن $Q_y = \left(\frac{y-\mu}{\sigma}\right)^2$.

$$X = U + V \sim WE(\alpha, \lambda).$$

تابع توزیع و تابع مولد گشتاور این توزیع به

ترتیب برابر با

$$F_X(x; \alpha, \lambda) = \frac{\alpha+1}{\alpha} [1 - e^{-\lambda x} - \frac{\lambda}{\lambda+\alpha\lambda} (1 - e^{-\lambda(1+\alpha)x})]; x > 0,$$

نکته در این مقاله برآوردهای ماکسیمم درستنمایی

توزیع t -چوله مورد استفاده قرار خواهند گرفت.

برای یافتن این برآوردها از تابع $st.mle$ از بسته

sn نرم افزار R استفاده می کنیم (آزایی و کاپیتانیو ۲۰۰۳) را ببینید.

۳ توزیع نمایی وزنی

تعریف ۱.۳. اگر X دارای تابع چگالی

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} \times \left(1 - \frac{t}{\lambda(1+\alpha)}\right)^{-1}; |t| < 1/\lambda$$

$$f_X(x; \alpha, \lambda) = \frac{\alpha+1}{\alpha} \lambda e^{-\lambda x} (1 - e^{-\alpha\lambda x}); x > 0$$

هستند.

باشد، گوییم متغیر تصادفی X دارای توزیع نمایی

وزنی با پارامترهای شکل $\alpha > 0$ و مقیاس $\lambda > 0$

است و آن را با نماد $WE(\alpha, \lambda)$ نشان می دهیم.

در حالت $\lambda = 1$ توزیع را با $WE(\alpha)$ نشان

می دهیم.

برای تولید اعداد تصادفی از توزیع نمایی وزنی

برآورد پارامترها از روش ماکسیمم درستنمایی و از

از لم زیر استفاده می شود.

روند زیر محاسبه می شوند.

لم ۲.۳. با فرض

براساس نمونه مشاهده شده $\{x_1, \dots, x_n\}$ و با استفاده

از تغییر پارامتر $\beta = \alpha\lambda$ تابع درستنمایی به صورت

زیر است:

$$U \sim \exp(\lambda),$$

$$V \sim \exp(\lambda(1+\alpha))$$

۴ تحلیل داده‌های واقعی و این تابع زیان به صورت زیر می‌باشد:

شبیه‌سازی

$$l(x_1, \dots, x_n) \quad (۱)$$

$$\begin{aligned} &= n \ln(\beta + \lambda) - n \ln \beta \\ &\quad + n \ln \lambda - \lambda \sum_{i=1}^n x_i \\ &\quad + \sum_{i=1}^n \ln(1 - e^{-\beta x_i}) \end{aligned}$$

در این بخش می‌خواهیم برتری توزیع t -چوله بر نمایی وزنی را نشان دهیم. برای این منظور در بخش (۱.۴) از داده‌های واقعی و در بخش (۲.۴) از

با فرض معلوم بودن β و با مشتق‌گیری از معادله داده‌های شبیه‌سازی شده استفاده می‌گردد.

(۱) نسبت به λ داریم:

برای مقایسه چند مدل، ملاک‌های متنوعی وجود

دارد که محققان بسته به سلیقه خود از آن‌ها استفاده

می‌کنند. این ملاک‌ها عبارتند از لگاریتم درستنمایی،

AIC ، BIC ، p -مقدار آزمون نیکویی برازش.

گوپتا و کاندو (۲۰۰۹) ملاک p -مقدار را بر دیگر

ملاک‌ها ترجیح داده‌اند. البته ارقامی و همکاران

(۲۰۰۸) معایب p -مقدار را به عنوان معیار پشتیبانی

داده‌ها از فرضیه صفر، مورد بررسی قرار داده‌اند.

سپس پیشنهاد می‌کنند که برای این منظور بهتر است

از نسبت درستنمایی مشاهدات تحت فرض‌های صفر

و مقابل استفاده شود. معیار پیشنهاد شده، معادل

مقایسه لگاریتم درستنمایی مشاهدات تحت فرض

صفر و فرض مقابل است. در هر حال برای مقایسه

دو مدل از همه ملاک‌های موجود استفاده می‌کنیم.

البته ملاک‌های شهودی مانند رسم هیستوگرام به

همراه چگالی متناظر برآورد شده و یا تابع بقا نیز

موجودند.

$$\hat{\lambda}(\beta) \quad (۲)$$

$$= \frac{1}{2\bar{x}} (\sqrt{(\beta\bar{x} - 2)^2 + 4\beta\bar{x}} - (\beta\bar{x} - 2))$$

که تابع درستنمایی را ماکسیم می‌کند. با جایگزینی

$\hat{\lambda}(\beta)$ در معادله (۱) داریم:

$$l(x_1, \dots, x_n) = n \ln(\beta + \hat{\lambda}(\beta)) \quad (۳)$$

$$- n \ln \beta + n \ln \hat{\lambda}(\beta)$$

$$- \hat{\lambda}(\beta) \sum_{i=1}^n x_i$$

$$+ \sum_{i=1}^n \ln(1 - e^{-\beta x_i})$$

معادله (۳) فقط شامل مقدار نامعلوم β است. پس

با استفاده از یک تابع اپتیمم، از ماکسیم نمودن

تابع $l(\beta)$ ، $\hat{\beta}$ را می‌یابیم. اکنون با استفاده از معادله

(۲) می‌توان $\hat{\lambda}$ را محاسبه کرد. سرانجام چون $\beta =$

$$\alpha \lambda \text{ پس } \hat{\alpha} = \frac{\hat{\beta}}{\hat{\lambda}}$$

۱.۴ داده‌های واقعی

چوله به صورت

داده‌های واقعی این بخش شامل دو مجموعه داده می‌باشد که گوپتا و کاندو (۲۰۰۹) برای نشان دادن برتری توزیع نمایی وزنی بر توزیع‌های گاما، وایبل و نمایی تعمیم یافته مورد استفاده قرار داده اند. در این بخش تنها نتایج مقایسه دوم مدل t -چوله و نمایی وزنی را بررسی می‌کنیم. برای مقایسه مدل نمایی وزنی با مدل‌های گاما، وایبل و نمایی تعمیم یافته می‌توانید بخش (۵) گوپتا و کاندو (۲۰۰۹) را ببینید.

$$\hat{\mu} = 31.927373, \hat{\sigma} = 52.392291, \\ \hat{\lambda} = 3.629520, \hat{\nu} = 2.018602 \\ \text{و برای توزیع نمایی وزنی برابر } \hat{\alpha} = 1.6232 \text{ و } \hat{\lambda} = 0.0138 \text{ به دست می‌آیند.}$$

شکل (۱) هیستوگرام داده‌ها به همراه توابع چگالی برازش داده شده به آن‌ها را نشان می‌دهد. با ملاحظه این نمودارها به طور شهودی توزیع t -چوله برازش مناسبتری به داده‌ها را پیشنهاد می‌کند (منحنی خط چین).

مثال ۱.۴. این داده‌ها زمان زنده ماندن خوکچه‌ها بعد از تزریق مقادیر مختلف باسیل سل بر حسب روز است که توسط بیجر کدال (۱۹۶۰) مورد مطالعه قرار گرفت. داده‌ها به شرح زیر است:

اکنون با معیارهای کمی به مقایسه دو مدل می‌پردازیم. یکی از معروف ترین ملاک‌ها لگاریتم درستنمایی است. اما از آنجا که تعداد پارامتر برآورد شده دو مدل یکسان نیست این ملاک چندان مناسب نیست. گرچه برخی از نویسندگان به این موضوع توجهی ندارند (برای مثال آرلانو-واله و همکاران (۲۰۰۳) را ببینید). به طور معمول برای مقایسه مدل‌هایی با تعداد پارامترهای متفاوت از ملاک اطلاع آکائیک (AIC) و ملاک اطلاع بیزی (BIC) استفاده می‌شود (لروکس (۱۹۹۲) را ببینید). معمولاً مدل با AIC کمتر بهتر است. این ملاکها از فرمول‌های $AIC = -2\log(L) + 2k$ و $BIC = -2\log(L) + k\log(N)$ محاسبه می‌شوند، که در آن‌ها L ماکسیمم درستنمایی، k تعداد پارامترهای مدل و N تعداد مشاهدات است. یکی دیگر از ملاک‌های عددی برآورد ماکسیمم درستنمایی پارامترهای توزیع t -

$x = \{12, 15, 22, 24, 24, 32, 33, 34, 38, 38, 43, 44, 48, 52, 53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60, 60, 76, 61, 62, 63, 65, 65, 67, 68, 70, 70, 72, 73, 75, 76, 81, 83, 84, 85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127, 129, 131, 143, 146, 146, 175, 175, 211, 233, 258, 258, 263, 297, 341, 341, 376\}$

p -مقدار آزمون نیکویی برازش است. چون p - دیگری متفاوت است)، دانشجوی به رشته پایین تر مقدار بیشتر، توجه بیشتری برای فرض صفر در منتقل می‌شود. در سال ۲۰۰۳، نمره آزمون نهایی مقابل فرض مقابل قرار دارد انتظار داریم که در دانشجویان منتقل شده به رشته پایین تر در درس یک آزمون نیکویی برازش، مدل مناسب تر p - مقدار بیشتری داشته باشد. از مهمترین آزمون‌های نیکویی برازش آزمون کلموگروف-اسمیرنف است. از این رو معمولاً p -مقدار این آزمون را با هم مقایسه می‌کنند. از طرفی چون مقدار آماره آزمون کلموگروف-اسمیرنف $(K - S)$ میزان فاصله بین تابع توزیع تجربی و برازش است، هرچه مقدار $K - S$ کوچکتر باشد، مدل مناسب تر است. توجه کنید که ملاک p -مقدار با $K - S$ معادل است. جدول (۱) خلاصه ای از ملاک‌های مقایسه دو مدل را برای داده‌ها نشان می‌دهد. با استفاده از هر یک از ملاک‌ها، مدل t -چوله بهتر است.

در این مورد $n = 48$ است. برای توزیع t -چوله برآوردهای

$$\hat{\mu} = 3.99999, \hat{\sigma} = 24.26756,$$

$$\hat{\lambda} = 2.208673, \hat{\nu} = 6.712828$$

مثال ۲.۴. موسسه فناوری کانپور هند، یکی از موسسات مشهور فنی پذیرش در رشته‌های مختلف است. این موسسه هر سال ۱ طریق انجام آزمون ورودی مشترک حدود ۴۵۰ دانشجوی می‌پذیرد. از آنجا که این پذیرش بسیار دشوار و رقابت نیز بسیار بالاست، پذیرش را به شیوه زیر انجام می‌دهند. در سال اول، دانشجویان همه رشته‌ها، دروس اجباری رشته خود را می‌گذرانند. در اولین آزمون اواسط ترم اگر دانشجویی نمره زیر نمره حد نصاب کسب کند (که این نمره در هر سال و از یک رشته به

و برآورد ماکسیمم درستنمایی پارامترهای توزیع $WE(\alpha, \lambda)$ برابر $\hat{\alpha} = 0.2919$ و $\hat{\lambda} = 0.0685$ به دست می‌آید. هیستوگرام داده‌ها به همراه توابع چگالی برازش داده شده به آن‌ها در شکل (۲) رسم شده‌اند. با توجه به نتایج جدول (۲)، در این مثال نیز توزیع t -چوله برتر است.

۲.۴ شبیه سازی

در این بخش می‌خواهیم با استفاده از داده‌های شبیه سازی شده کارائی توزیع t -چوله را نشان دهیم.

برای این منظور نمونه‌های تصادفی به اندازه‌های تراست.

$n = \{50, 100\}$ از توزیع‌های وایبل، گاما و نمایی

۵ نتیجه گیری

تعمیم یافته (GE) با پارامترهای مختلف تولید کرده

و سپس آن‌ها را با استفاده از توزیع‌های t -چوله

و نمایی وزنی برازش می‌دهیم. دلیل انتخاب این

توزیع‌ها این است که مدل نمایی وزنی بر این مدل‌ها

ارجحیت داد (گوپتا و کاندو (۲۰۰۹)). ملاک‌های

مقایسه دو مدل همانند بخش قبل به کار می‌روند.

جداول ۳ و ۴ نشان می‌دهند که در اکثر موارد،

توزیع t -چوله برازش بهتری به داده‌ها داد. شکل

۳، هیستوگرام نمونه‌های به اندازه $n = 50$ به همراه

چگالی‌های برازش شده را نشان می‌دهد.

چون در برخی از موارد نتایج خیلی به هم نزدیک

به دست آمده‌اند، لذا ناچار شدیم که از تعداد نمونه‌های

تصادفی بیشتری استفاده کنیم و درصد مواردی که

بر اساس هر یک از ملاک‌ها، مدل t -چوله برتری

ارد را گزارش دهیم. به این منظور هر بار با استفاده

از $B = 100$ نمونه به اندازه $n = \{50, 100\}$ ،

درصد مواردی که در آن‌ها توزیع t -چوله بهتر است را

محاسبه نمودیم. نتایج در جدول (۵) گزارش شده

اند. ملاحظه می‌شود که در همه موارد در بیش از

نیمی از نمونه‌ها p -مقدار مدل t -چوله بیشتر است.

همچنین در ۷ مورد از ۱۲ مورد بررسی شده، AIC

بیش از نیمی از موارد برای این مدل کمتر است. با

استفاده از این نتایج برتری توزیع t -چوله توجیه پذیر

در این مقاله با استفاده از مثال‌هایی از داده‌های واقعی و همچنین شبیه سازی نشان دادیم که مدل t -چوله جایگزین مناسبی برای توزیع نمایی وزنی است. برای مقایسه دو مدل از معیارهای مختلف استفاده شد. توجه کنید که مدل t -چوله بر توزیع‌های گاما، وایبل و نمایی تعمیم یافته نیز اولویت دارد. زیرا گوپتا و کاندو (۲۰۰۹) برتری مدل نمایی وزنی را بر توزیع‌های گاما، وایبل و نمایی تعمیم یافته نشان داده‌اند.

جدول ۱: مقایسه مدل‌های t -چوله و نمایی وزنی برای برازش داده‌ها مثال (۱).

مدل	$K - S$	$P - value$	$\log - likelihood$	AIC
WE	0.1173	0.2748	-393.5689	791.1379
t -چوله	0.0854	0.6701	-390.8064	789.6427

جدول ۲: مقایسه مدل‌های t -چوله و نمایی وزنی برای برازش داده‌ها مثال (۲) با ملاک‌های مختلف. ه.

مدل	$K - S$	$P - value$	$\log - likelihood$	AIC
WE	0.0933	0.7978	-197.3370	398.6740
t -چوله	0.0854	0.8752	-195.3052	398.6105

جدول ۳: مقایسه توزیع‌های t -چوله و نمایی وزنی برای نمونه‌های شبیه سازی شده به اندازه $n = 50$

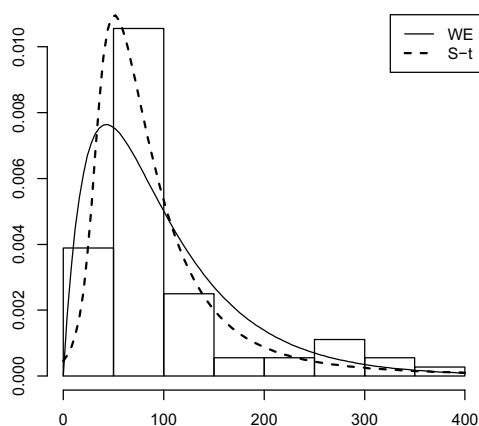
مدل	$Weibull(0.5, 1)$			$Weibull(2, 3)$		
	$K - S$	$P - value$	AIC	$K - S$	$P - value$	AIC
WE	0.2788	0.0006278	123.3373	0.1255	0.3784	176.0947
t -چوله	0.1413	0.2467	114.413	0.0599	0.9891	172.281
مدل	$Gamma(0.5, 1)$			$Gamma(2, 4)$		
	$K - S$	$P - value$	AIC	$K - S$	$P - value$	AIC
WE	0.204	0.02663	59.55727	0.0873	0.8086	18.43670
t -چوله	0.196	0.03696	63.9841	0.0783	0.8955	16.97318
مدل	$GE(0.5, 1)$			$GE(1.5, 2)$		
	$K - S$	$P - value$	AIC	$K - S$	$P - value$	AIC
WE	0.182	0.036378	38.46457	0.159	0.1431	50.57952
t -چوله	0.1805	0.06748	43.72241	0.0984	0.682	51.58882

جدول ۴: مقایسه مدل‌های t -چوله و نمایی وزنی برای نمونه‌های شبیه سازی شده به اندازه $n = 100$

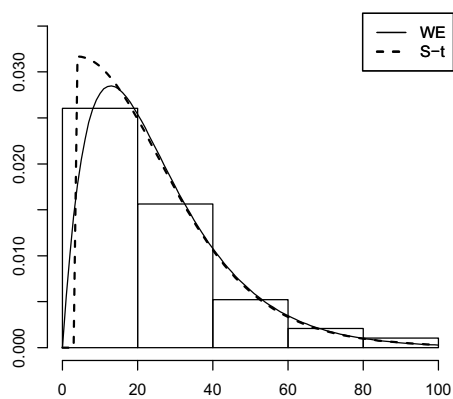
	<i>Weibull</i> (0.5, 1)			<i>Weibull</i> (2, 3)		
مدل	$K - S$	$P - value$	AIC	$K - S$	$P - value$	AIC
<i>WE</i>	0.2808	$2.847e - 07$	353.543	0.1448	0.03020	367.5081
t -چوله	0.1221	0.1015	321.3264	0.0745	0.6357	357.751
	<i>Gamma</i> (0.5, 1)			<i>Gamma</i> (2, 4)		
مدل	$K - S$	$P - value$	AIC	$K - S$	$P - value$	AIC
<i>WE</i>	0.1718	0.005472	119.0600	0.0678	0.7473	41.02609
t -چوله	0.1256	0.08523	114.8250	0.052	0.9499	41.06022
	<i>GE</i> (0.5, 1)			<i>GE</i> (1.5, 2)		
مدل	$K - S$	$P - value$	AIC	$K - S$	$P - value$	AIC
<i>WE</i>	0.1641	0.009188	111.4069	0.0602	0.8616	116.7340
t -چوله	0.1333	0.0571	110.9134	0.0834	0.4893	117.2572

جدول ۵: درصد مواردی که در $B = 100$ نمونه، p -مقدار مدل t -چوله بیشتر و AIC آن کمتر است.

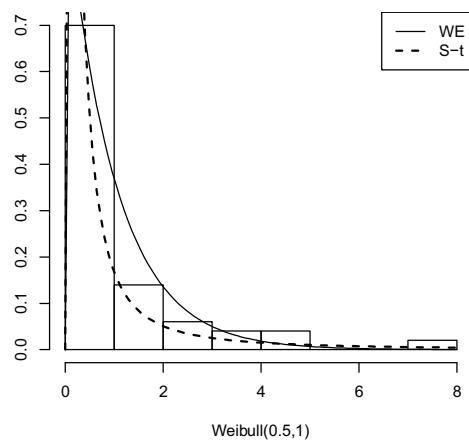
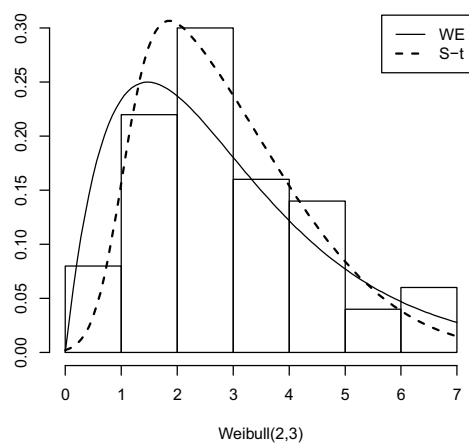
<i>Weibull</i> (0.5, 1)			<i>Weibull</i> (2, 3)		
n	$P - value$	AIC	n	$P - value$	AIC
$n = 50$	1	0.97	$n = 50$	0.95	0.87
$n = 100$	0.96	0.94	$n = 100$	0.96	0.96
<i>Gamma</i> (0.5, 1)			<i>Gamma</i> (2, 4)		
n	$P - value$	AIC	n	$P - value$	AIC
$n = 50$	0.89	0.53	$n = 50$	0.53	0.35
$n = 100$	0.97	0.73	$n = 100$	0.55	0.35
<i>GE</i> (0.5, 1)			<i>GE</i> (1.5, 2)		
n	$P - value$	AIC	n	$P - value$	AIC
$n = 50$	0.86	0.31	$n = 50$	0.65	0.36
$n = 100$	0.95	0.55	$n = 100$	0.55	0.31

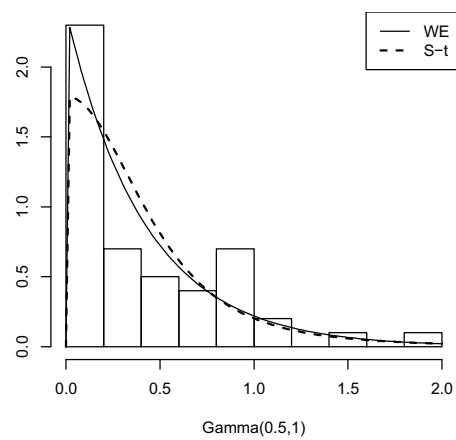
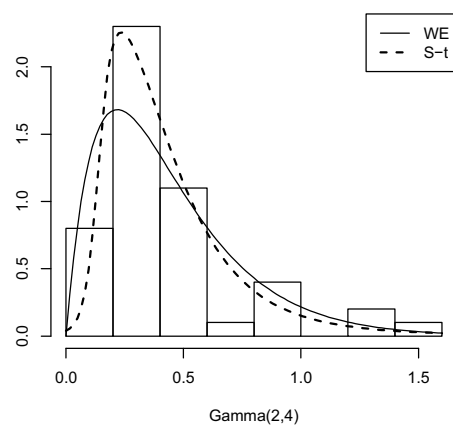


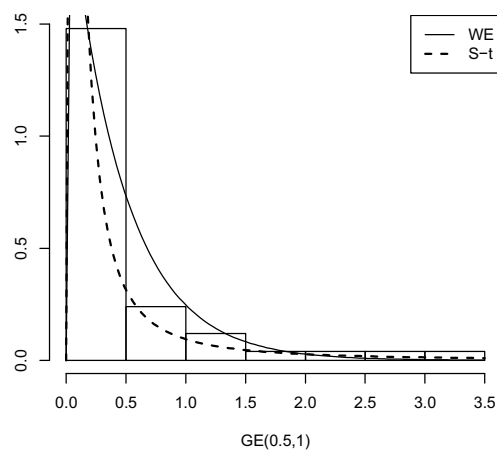
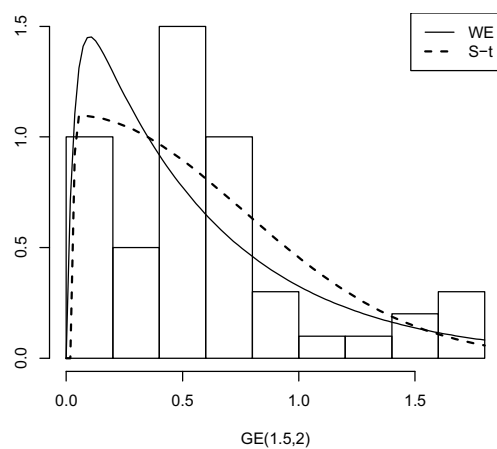
شکل ۳: هیستوگرام داده‌های مثال ۱.۴ به همراه چگالی‌های برازش شده به آن.



شکل ۴: هیستوگرام داده‌های مثال ۲.۴ به همراه چگالی‌های برازش شده به آن.







مراجع

- [۱] ارقامی، ن. ر.، زمان زاده، ا. و راحت، س. (۱۳۸۷)، شواهد آماری، قسمت دوم: قانون درستنمایی و معایب p -مقدار به عنوان معیار پشتیبانی داده‌ها از فرضیه صفر، اندیشه آماری، سال ۱۳، شماره ۲، ۱۵-۳.
- [2] Arellano-Valle, R.B., Gomez, H.W., Quintana, F.A. (2003). A new class of skew normal distributions. *Communications in Statistics-Theory and Methods*, 33(7), 1465-1480.
- [3] Azzalini, A. (1985). A class of distribution which includes the normal ones, *Scandinavian Journal of Statistics*, 12, 171-178.
- [4] Azzalini, A. and Capitanio, A. (2003). Distribution generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc. Ser. B* 65, 367-389.
- [5] Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli, *Amer. J. Hyg*, 72, 130-148.
- [6] Gupta, R. D. and Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments, *Journal of Statistical planning and Inference*, 137, 173-188.
- [7] Gupta, R. D. and Kundu, D. (2009). A new class of weighted exponential distributions, *Statistics*, 43, 621-643.
- [8] Leroux, B. G. (2007/1992). Consistent estimation of a mixing distribution, *Annals of Statistics*, 20 (3), 1350-1360.