بررسی توزیع بتا- نرمال، مقایسه با گاما - نرمال و سایر تلفیق‌های توزیع بتا

فتانه نظام پور،[۱] علیرضا سلیمانی[۲]

تاریخ دریافت: ۱۳۹۶/۷/۶
تاریخ پذیرش: ۱۳۹۷/۴/۲۷

چکیده:
در این مقاله خانوادهٔ بتا - X به‌صورت کلّی و عضوی از این خانواده، توزیع بتا - نرمال، در جزئیات مورد مطالعه قرار گرفته است. از یک مجموعه‌داده واقعی برای نشان دادن کاربرد توزیع بتا - نرمال و همچنین مقایسهٔ این توزیع با توزیع‌های گاما - نرمال و بیرن بام سندرس، استفاده شده است.

واژه‌های کلیدی: توزیع بتا، توزیع بیرن بام سندرس، توزیع گاما - نرمال، معیار اطلاع آکائیکه، معیار اطلاع بیزی، برآورد ماکسیمم درست‌نمایی.

۱ مقدمه

یکی ازویژگی‌های توزیع‌های تعمیم‌یافته بتا توانایی آنها در برازش داده‌های چوله است که به‌طور صریح توسط توزیع‌های موجود نمی‌تواند برازش داده شود [۱]. در ابتدا کلاسی از توزیع‌های بتای تعمیم‌یافته استفاده از رابطهٔ زیر تعریف می‌شود:

$$G(x) = \frac{1}{\beta(a,b)} \int_0^{F(x)} t^{a-1}(1-t)^{b-1}dt \qquad (1)$$

چند توزیع بتا تعمیم‌یافته در سال‌های اخیر مورد بحث قرار گرفته است. [۳، ۴، ۵] به‌ترتیب توزیع‌های بتا - نرمال، بتا - گامبل، بتا-فرشت و بتا - نمایی را پیشنهاد کردند. تابع چگالی متناظر با (۱) به‌صورت زیر است:

$$g(x) = \frac{f(x)}{\beta(a,b)}(F(x))^{a-1}(1-F(x))^{b-1} \qquad (2)$$

تابع چگالی توزیع بتا - نرمال (۲) با نماد $BN(\mu,\sigma,a,b)$ نمایش داده می‌شود و به‌صورت عبارت زیر است:

$$g(x) = \frac{1}{\beta(a,b)}(\Phi(\frac{x-\mu}{\sigma}))^{a-1}(1-\Phi(\frac{x-\mu}{\sigma}))^{b-1} \times \phi(\frac{x-\mu}{\sigma})\sigma^{-1}. \qquad (3)$$

در بخش‌های بعد توزیع بتا - نرمال با توزیع‌های بتا - بیرن بام سندرس و بیرن بام سندرس و توزیع گاما - نرمال مقایسه می‌شوند و نشان داده می‌شود که توزیع بتا - نرمال برای مدل‌بندی واقعی الیاف کربن (داده‌های چوله) از بقیه توزیع‌ها عمل می‌کند. برای این کار از نرم‌افزارهای SAS و R استفاده کرده‌ایم.

۲ توزیع بتا - نرمال: مشخصات و ویژگی‌ها

چندین روش برای تولید توزیع‌های پیوسته وجود دارد. بسیاری از این روش‌ها در کتاب [۱۱] مورد بحث قرار گرفته است. پس از انتشار این کتاب، روش‌های جدیدی در تحقیقات ظاهر شد. [۲] کلاسی از توزیع‌های تولید شدهٔ بتا را معرفی کردند و اشاره نمودند که توزیع آماره‌های ترتیبی حالات خاصی از توزیع‌های تولید شدهٔ بتا است.

[۱۲] برخی از خواص توزیع‌های تولید شدهٔ بتا را مورد مطالعه قرار داد. بسیاری از توزیع‌های تولید شدهٔ بتا قبلاً مطالعه شده‌اند (به‌عنوان مثال: [۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸] را ببینید).

[۱] دانش‌آموختهٔ کارشناسی ارشد آمار ریاضی، دانشگاه حکیم سبزواری، ایران.
[۲] عضو هیئت علمی گروه آمار، دانشگاه حکیم سبزواری، ایران.

روشی که منجر به توزیع‌های تولید شده بتا شده است با استفاده از یک توزیع بتا تعمیم‌یافته به‌عنوان مولّد گسترش یافت [۱۹، ۲۰]. [۲۱]روش تبدیل انتگرال احتمال معکوس را برای توزیع‌های چوله به کار بردند، که شامل خانواده نرمال چوله معرفی‌شده توسط آزالینی [۲۲، ۲۳]، به‌عنوان یک کلاس خاص بود.

اگر V دارای توزیع بتا با پارامترهای a و b باشد، آن‌گاه $x = \sigma\Phi^{-1}(V) + \mu$ دارای توزیع بتا - نرمال با پارامترهای μ و σ و a و b است که تابع چگالی احتمال آن به‌صورت (۳) است. هنگامی که $a = b = 1$، توزیع بتا - نرمال به توزیع نرمال تقلیل می‌یابد که توزیع بتا - نرمال به مراتب از توزیع نرمال جذاب‌تر است. می‌توان تابع نرخ خطر توزیع بتا - نرمال را به شکل زیر بیان نمود:

$$h(x) = \frac{1}{\beta(a,b)}\left(\Phi(\frac{x-\mu}{\sigma})\right)^{a-1}\left(1-\Phi(\frac{x-\mu}{\sigma})\right)^{b-1} \times \phi(\frac{x-\mu}{\sigma})\sigma^{-1}\left(1-I_{\Phi(\frac{x-\mu}{\sigma})}(a,b)\right)^{-1}. \quad (۴)$$

با توجه به تعریف تابع چگالی احتمال توزیع بتا - نرمال، می‌توان تابع چگالی را به‌صورت یک سری نامتناهی نمایش داد، که به‌صورت زیر است:

$$g(x) = \sum_{j=0}^{\infty}(-1)^j \frac{(b-1)...(b-j)}{\sigma j!}\phi(\frac{x-\mu}{\sigma}) \times (\Phi(\frac{x-\mu}{\sigma}))^{a+j-1}. \quad (۵)$$

همچنین می‌توان متمم تابع توزیع بتا - نرمال را به‌صورت یک سری نامتناهی به‌صورت زیر محاسبه نمود:

$$1 - G(x) = \sum_{j=0}^{\infty}(-1)^j \frac{(b-1)...(b-j)}{(a+j)j!} \times (1-\Phi(\frac{x-\mu}{\sigma}))^{a+j}. \quad (۶)$$

و در نهایت می‌توان تابع نرخ خطر در عبارت (۴) را به‌صورت تقسیم دو سری نوشت:

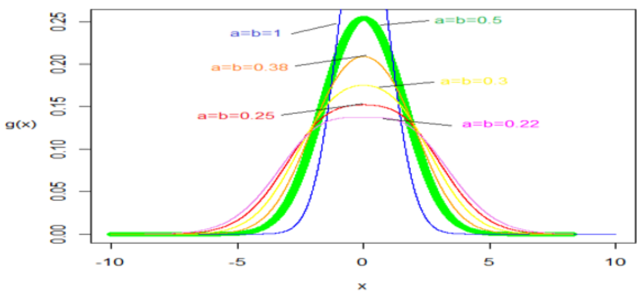
$$h(x) = \frac{\sum_{j=0}^{\infty}(-1)^j \frac{(b-1)...(b-j)}{\sigma j!}\phi(\frac{x-\mu}{\sigma})(\Phi(\frac{x-\mu}{\sigma}))^{a+j-1}}{\sum_{j=0}^{\infty}(-1)^j \frac{(b-1)...(b-j)}{(a+j)j!}(1-\Phi(\frac{x-\mu}{\sigma}))^{a+j}}. \quad (۷)$$

r-امین گشتاور غیرمرکزی توزیع بتا - نرمال به‌صورت زیر است:

$$E(x^r) = \frac{1}{\beta(a,b)}\int_0^1 u^{a-1}(1-u)^{b-1}(\sigma\Phi^{-1}(u)+\mu)^r du \quad (۸)$$

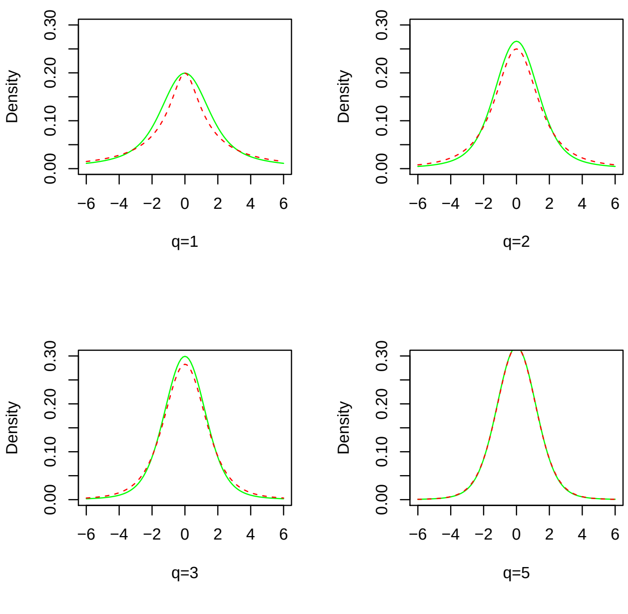
که با جایگذاری $u = \Phi(\frac{x-\mu}{\sigma})$ در عبارت (۴) نتیجه به دست می‌آید.

اکنون چند ویژگی از چگالی توزیع بتا - نرمال با استفاده از نمودارها (کلیّهٔ نمودارها با کمک نرم‌افزار R رسم شده‌اند) بررسی می‌شود. شکل ۱ نشان می‌دهد که اگر $a = b$، توزیع بتا - نرمال بر روی μ متقارن است.



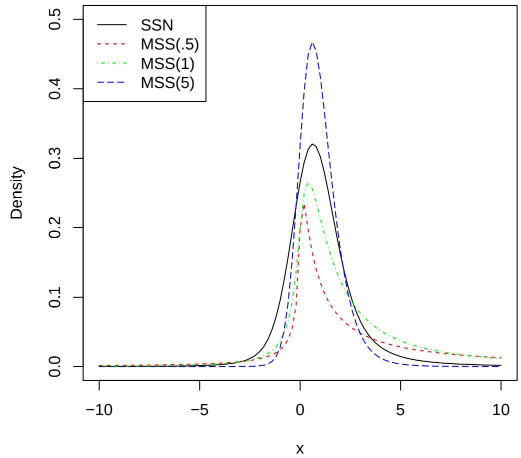
شکل ۱. توزیع بتا - نرمال $BN(0,1,a,b)$ وقتی $a = b$

با توجه به شکل ۲، توزیع بتا - نرمال وقتی $a > b$ به راست چوله است که با افزایش b، چولگی راست کاهش می‌یابد.



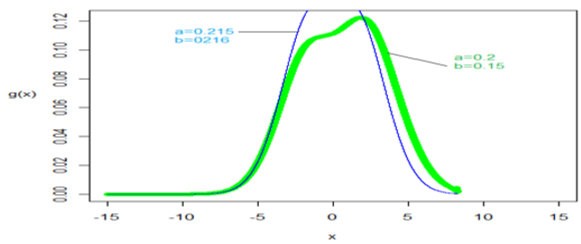
شکل ۲. توزیع بتا - نرمال $BN(0,1,0/5,b)$ به‌ازای bهای مختلف

با توجه به شکل ۳، توزیع بتا - نرمال وقتی $a < b$ به چپ چوله است که با افزایش a، چولگی چپ کاهش می‌یابد.



شکل۳. توزیع بتا - نرمال $BN(\circ,1,a,\circ/5)$ به‌ازای aهای مختلف

اگر $a > 1$ و $b > 1$، آن‌گاه توزیع کشیده می‌شود و اگر $a < 1$، $b < 1$ و $a = b$ توزیع به‌صورت متقارن و پهن است [۲]. در توزیع بتا - نرمال اگر a و b از ۰٫۲۱۴ بزرگ‌تر و یا مساوی باشند، آن‌گاه توزیع تک نمایی است و اگر a و b از ۰٫۲۱۴ کوچک‌تر باشند، آن‌گاه توزیع دونمایی است [۱۰]، این موضوع به وضوح در شکل ۴ پیداست.



شکل۴. توزیع بتا - نرمال در حالت تک نمایی و دو نمایی

۳ توزیع گاما - نرمال

[۲۴] روش جدیدی برای تولید خانواده‌ای از توزیع‌ها توسعه دادند و آن را خانوادهٔ $T-X$ از توزیع‌ها نامیدند. برای بررسی روش‌ها جهت تولید توزیع‌های پیوسته یک متغیره، به [۲۵] مراجعه شود. در ابتدا، T به‌عنوان یک متغیر تصادفی گاما و X به‌عنوان یک متغیر تصادفی پیوسته در نظر گرفته شود برخی خواص کلّی از خانوادهٔ گاما-X را مطالعه می‌شود. سپس، جزئیات توزیع گاما - نرمال که یک عنصر از خانوادهٔ گاما -X است، مطالعه می‌شود.

جدول ۱. فشار بر اثر شکستن الیاف کربن به طول ۵۰ میلی متر

۲٫۸۲	۳٫۲۲	۲٫۵۰
۳٫۱۹	۲٫۷۳	۲٫۵۹
۲٫۸۸	۱٫۸۴	۳٫۲۲
۲٫۴۱	۲٫۷۴	۲٫۵۵
۳٫۷۰	۴٫۳۸	۲٫۹۳
۱٫۲۵	۲٫۶۷	۴٫۴۲
۳٫۲۷	۲٫۳۵	۱٫۸۹
۰٫۸۵	۳٫۳۳	۳٫۰۹
۴٫۲۰	۳٫۲۸	۳٫۱۱
۲٫۱۲	۳٫۷۵	۲٫۴۸
۲٫۸۱	۱٫۶۹	۳٫۶۰
۱٫۶۱	۳٫۶۵	۳٫۶۸
۲٫۰۵	۰٫۳۹	۳٫۳۹
۴٫۹۰	۳٫۱۱	۲٫۰۳
۳٫۳۹	۴٫۷۰	۳٫۳۱
۲٫۷۹	۳٫۳۱	۱٫۸۰
۱٫۴۷	۲٫۵۶	۲٫۹۷
۱٫۸۷	۲٫۸۷	۳٫۱۵
۲٫۹۵	۱٫۶۱	۲٫۵۵
۳٫۱۵	۲٫۴۳	۱٫۰۸
۲٫۸۵	۱٫۵۷	۳٫۵۶
۲٫۵۳	۲٫۹۶	۲٫۰۳

فرض کنید $F(x)$ تابع توزیع تجمعی متغیر تصادفی X و $r(t)$ تابع چگالی احتمال متغیر تصادفی T تعریف شده روی $[\circ,\infty)$ باشد. [۲۴] تابع توزیع تجمعی خانواده $T-X$ از توزیع‌های تعریف‌شده به‌صورت زیر به دست می‌آید:

$$G(x) = \int_{\circ}^{-log(1-F(x))} r(t)dt. \qquad (9)$$

[۲۴] خانواده توزیع‌های تعریف‌شده در (۹) را خانوادهٔ مبدل - تبدیل‌شده (یا خانواده $T-X$) نامیدند. هنگامی که X یک متغیر تصادفی پیوسته است، تابع چگالی احتمال خانواده $T-X$ به‌صورت زیر است:

$$g(x) = \frac{f(x)}{1-F(x)} r(-log(1-F(x))) = h(x)r(H(x)), \qquad (10)$$

بنا بر این خانواده توزیع‌های تعریف‌شده در(۱۰) می‌توانند به‌عنوان یک خانواده توزیع‌های ناشی‌شده از توابع خطر، در نظر

گرفته شوند.

اگر متغیر تصادفی T از توزیع گاما با پارامترهای α و β و متغیر X دارای توزیع نرمال با پارامترهای μ و σ^2 باشد، آن‌گاه تعریف (۱۰) تبدیل به خانوادهٔ گاما - نرمال با تابع چگالی احتمال زیر می‌شود:

$$g(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\phi(x)(-log(1-\Phi(x)))^{a-1} \times (1-\Phi(x))^{1/\beta-1} \quad (11)$$

اگر متغیر تصادفی Y دارای توزیع گاما با پارامترهای α و β باشد، آن‌گاه متغیر تصادفی $X = \Phi^{-1}(1-e^{-Y})$ دارای توزیع گاما - نرمال با پارامترهای α و β و μ و σ^2 است.

داده‌های نمونهٔ تصادفی از توزیع گاما - نرمال در (۱۱) می‌توانند با اوّلین نمونهٔ تصادفی شبیه‌سازی شده Y، از توزیع گاما با پارامترهای α و β و پس از آن $X = \Phi^{-1}(1-e^{-Y})$ از توزیع گاما - نرمال شبیه‌سازی شوند [۸].

۴ کاربردها

در این بخش، یک مجموعه‌داده واقعی به‌منظور بررسی چگونگی عملکرد روش‌های بیان‌شده درنظر گرفته می‌شود. داده‌ها در جدول ۱ فشار بر اثر شکستن الیاف کربن به طول ۵۰ میلی متر [۶] داده شده است.

[۷] یک توزیع چهار پارامتری بتا - بیرن بام سندرس (BBS) را پشنهاد کردند. همچنین [۸] توزیع گاما - نرمال را پیشنهاد دادند.

برآوردهای ماکسیمم درست‌نمایی، مقدار لگ درست‌نمایی، AIC (معیار اطلاع آکائیکه) و BIC (معیار اطلاع بیزی) توزیع‌های برازش داده‌شده در جدول ۲ گزارش شده‌اند و نتایج برازش مجموعه‌داده‌های توزیع بتا - نرمال را با توزیع‌های بیرن بام سندرس دو پارامتری [۹]، بتا - بیرن بام سندرس چهارپارامتری [۷] و گاما - نرمال [۸] مقایسه شده‌اند. کلیّه محاسبات با استفاده از دستور NLMixed در SAS و دستور nlminb در R انجام شده است.

همان‌طور که در جدول ۲ مشاهده می‌شود، معیار اطلاع آکائیکه و بیزی برای توزیع بتا - نرمال دارای کمترین مقدار بین سایر توزیع‌های مذکور است و این بدان معناست که توزیع بتا - نرمال برازش بهتری نسبت به توزیع‌های بیرن بام سندرس و بتا - بیرن بام سندرس و گاما - نرمال برای داده‌های الیاف کربن دارد.

جدول۲. برآورد پارامترها برای داده‌های الیاف کربن

توزیع	برآورد پارامترها	لگاریتم درست‌نمایی	AIC	BIC
بیرن بام سندرس	$\hat{a} = 0/4371$, $\hat{b} = 2/5154$	$-100/19$	۲۰۴/۳۸	۲۰۶/۵۶۹۷
بتا - بیرن بام سندرس	$\hat{\alpha} = 0/1930$, $\hat{\beta} = 1876/732$, $\hat{a} = 1/0445$, $\hat{b} = 57/6001$	$-91/3550$	۱۹۰/۷۱	۱۹۹/۴۶۸۶
گاما - نرمال	$\hat{\mu} = 3/3464$, $\hat{\sigma} = 0/9614$, $\hat{\alpha} = 0/8263$, $\hat{\beta} = 0/6190$	$-85/43896$	۱۷۸/۸۷۹۲	۱۸۷/۶۳۷۸
بتا - نرمال	$\hat{\mu} = 3/41067$, $\hat{\sigma} = 1/1826$, $\hat{a} = 1/1823$, $\hat{b} = 2/3913$	$-85/4389$	۱۷۸/۸۷۷۸	۱۸۷/۶۳۶۵

۵ نتیجه‌گیری

در این بررسی خانوادهٔ بتا - X معرفی و حالت خاصی از خانوادهٔ بتا - X، (توزیع بتا - نرمال) مطالعه شد. مجموعه‌دادهٔ واقعی از توزیع بتا - نرمال با دیگر توزیع‌های شناخته شده مقایسه شد. نتایج نشان می‌دهد توزیع بتا - نرمال برازش بهتری برای مجموعه‌داده‌های الیاف کربن ارائه می‌دهد. در نتیجه پیشنهاد می‌شود برای داده‌های واقعی الیاف کربن از توزیع بتا - نرمال استفاده شود.

مراجع

[1] Castellares, F., Montenegro, L. C. and Cordeiro, G. M. (2010). The beta log-normal distribution. *Journal of Statistical Computation and Simulation*, **83**(**2**), 203-228.

[2] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications, Communications in Statistics, *Theory and Methods*, **31**(**4**), 497–512.

[3] Nadarajah, S. and Kotz, S. (2004). The beta Gumbel distribution, *Mathematical Probability*, **10**, 323-332.

[4] Nadarajah, S. and Gupta, A. K. (2004). The beta frechet distribution far east, *Journal of Theoretical Statistics*, **14**, 15-24.

[5] Nadarajah, S. and Kotz, S. (2005). The beta exponential distribution, *Reliability Engineering and System Safety*, **91**, 689-697.

[6] Nichols, M. D. and Padgett, W. J. (2006). A bootstrap control for Weibull percentiles,*Quality and Reliability Engineering International*, **22**, 141–151.

[7] Cordeiro, G. M. and Lemonte, A. J. (2011). The β-nirnbaum–saunders distribution: An improved distribution for fatigue life modeling, C*omputational Statistics and Data Analysis*, **55** (**3**), 1445–1461.

[8] Alzaatreh, A., Famoye, F. and Lee, Carl. (2014). The gamma-normal distribution: Properties and applications, *Computational Statistics and Data Analysis*, **69**, 67–80.

[9] Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions, *Journal of Applied Probability*, **6**, 319–327.

[10] Eugene, N. (2001). *A Generalized Normal Distribution: Properties, Estimation and Applications*, Unpublished Doctoral Dissertation: Central Michigan University, Mount Pleasant, Michigan.

[11] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, **1**, second ed. John Wiley and Sons, Inc., New York.

[12] Jones, M. C. (2004). Families of distributions arising from distributions of order statistics, **13**(**1**), 1–43.

[13] Famoye, F., Lee, C. and Eugene, N. (2004). Beta-normal distribution: bimodality properties and applications, *Journal of Modern Applied Statistical Methods*, **3**(**1**), 85–103.

[14] Famoye, F., Lee, C. and Olumolade, O. (2005). The beta-Weibull distribution, *Journal of Statistical Theory and Applications*, **4**(**2**), 121–136.

[15] Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution, *Reliability Engineering and System Safety*, **91**, 689–697.

[16] Akinsete, A. Famoye, F. Lee, C. (2008). The beta-Pareto distribution, *Statistics*, **42**, 547–563.

[17] Barreto-Souza, W., Santos, A. and Cordeiro, G. (2010). The beta generalized exponential distribution, *Journal of Statistical Computation and Simulation*, **80**(**2**),159–172.

[18] Alshawarbeh, E., Lee, C. and Famoye, F. (2012). Beta-Cauchy distribution, *Journal of Probability and Statistical Science*, **10**, 41–58.

[19] Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with tractability advantages, *Statistical Methodology*, **6**, 70–81.

[20] Cordeiro, G. M. and Castro, M. (2011). A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, **81**(**7**), 883–898.

[21] Ferreira, J. T. A. S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions, *Journal of the American Statistical Association*, **101**(**474**), 823–829..

[22] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[23] Azzalini, A. (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics*, **32**(**2**), 159–188.

[24] Alzaatreh, A., Lee, C. and Famoye, F. (2013b). A new method for generating families of continuous distributions, *Metron*, **71**(**1**), 63–79.

[25] Lee, C., Famoye, F. and Alzaatreh, A. (2013). Methods for generating families of univariate continuous distributions in the recent decades, *WIREs Computational Statistics*, **5**, 219–238.

[26] Alzaatreh, A., Lee, C. and Famoye, F. (2013b). A new method for generating families of continuous distributions, *Metron*, **71**(**1**), 63–79.

[27] Alzaatreh, A., Lee, C. and Famoye, F. (2012). On the discrete analogues of continuous distributions, *Statistical Methodology*, **9**, 589–603.

[28] Lee, C., Famoye, F. and Alzaatreh, A. (2013). Methods for generating families of univariate continuous distributions in the recent decades, *WIREs Computational Statistics*, **5**, 219–238.