إمتداد جديد لتوزيع Yule بثلاث معالم

إعداد

الأستاذ الدكتور/ محمد توفيق إسماعيل البلقيني

أستاذ الرياضيات والإحصاء الإكتواري

كلية التجارة – جامعة المنصورة

الأستاذ الدكتور/ البيومي عوض عوض طاقية

أستاذ بقسم الإحصاء التطبيقي والتأمين

كلية التجارة – جامعة المنصورة

الدكتور / رضا عبد الفتاح مصطفى الشناوي

مدرس بقسم الإحصاء التطبيقي والتأمين

كلية التجارة – جامعة المنصورة

أمل طلبة محمد طلبة بدوي

مدرس مساعد بقسم الرياضيات والإحصاء والتأمين

أكاديمية السادات للعلوم الإدارية

الملخص

إتجهت بعض الدراسات الحديثه إلى إشتقاق إمتدادات أخرى لتوزيع Yule والذي يعد من أكثر وأدق التوزيعات المستخدمه في البيانات شديدة الإلتواء وذلك بهدف الحصول على دقه أعلى للتوزيع، من أهم تلك الدراسات دراستي [Spierdijk (٢٠٠٧)] و [Rodríguez (٢٠١١)].

في هذا البحث تم الربط بين النتائج التي إنتهت إليها كلا الدراستين المشار إليهما ومن ثم تم الحصول على إمتداد آخر جديد لتوزيع [Yule (ρ)] بثلاثة معالم وذلك بإستخدام أسلوب خلط التوزيعات. كما تم الحصول على تفسير لتباين التوزيع من خلال تقسيمة إلى ثلاثة أجزاء بحسب مصادر الإختلاف (عشوائية – داخليه – خارجية).

التوزيع الجديد المقدم في هذا البحث يعتبر الأعم والأشمل لتوزيعات Yule حيث أنه يتضمن الإمتدادات المقدمة في الدراستين المشار لهما كحالات خاصة منه. وقد تم إستخدام التوزيع الجديد في توفيق نفس نوع البيانات المستخدم في دراستي [Spierdijk (٢٠٠٧)] و [Rodríguez (٢٠١١)] ومقارنة نتائج البحث مع النتائج التي تم الحصول عليها في تلك الدراسات.

ABSTRACT

Some recent studies have tended to derive other extensions for Yule, which is the more efficient one for fitting superstar data, to obtión more accuracy in fitting data. The two more important studies are [Spierdijk (٢٠٠٧), Rodríguez (٢٠١١)].

This research linked between the findings of both studies referred to, and then another new extension for Yule distribution with three parameters was given, using the method of mixture distributions. Then an explanation of the variance of the new distribution was given by partitioning it into three components: randomness, liability and proneness.

The new distribution is more general for Yule distributions, since it includes extensions provided in the two studies- that mentioned above as special cases. This research aims to find maximum likelihood estimators of the new distribution numerically for the same kind of data used in [Spierdijk (٢٠٠٧), Rodríguez (٢٠١١)], comparing the results with those obtained in these studies.

Key words: Yule distribution, generalized Yule distribution, extended Yule distribution, Appell hypergeometric function, Mixed distributions, superstar data.

مقدمة

قدم توزيع Yule(ρ) والمعروف بإسم توزيع Yule-Simon في دراسة [Yule, G. (١٩٢٥)] ثم أعيد إكتشافه في دراسة [Simon (١٩٥٥)]. تأخذ الدالة الإحتماليه للمتغير

العشوائي Y والذي يتبع توزيع $\text{Yule}(\rho)$ الشكل التالي:

$$P[Y=y]$$
$$=\rho B(y,\rho+١)$$
$$;\quad y=١,٢,\ldots$$

حيث أن ρ هي معلمة التوزيع الوحيدة $(\rho>٠)$ و $B(a,b)$ تمثل دالة بيتا.

إهتمت دراسة [Spierdijk (٢٠٠٧)] بما أشارت إليه دراستي [Yule (١٩٢٤, sec. II)] و [Simon (١٩٥٥, sec. I)] فيما يخص إمكانية إشتقاق تعميم آخر من توزيع Yule بمعلمتين. وقد تم الحصول على صيغة جديده لتوزيع Yule بمعلمتين وأطلق عليه GYule إختصارا لـ generalized Yule ، حيث تأخذ الدالة الإحتماليه للمتغير العشوائي Y والذي يتبع توزيع $\text{GYule}(\rho,\alpha)$ الشكل التالي:

$$P[Y=y]$$
$$=\frac{\rho}{١-\alpha^{\rho}}B_{١-\alpha}(y,\rho+١)$$
$$;\quad y=١,٢,\ldots$$

حيث أن:

- ρ و α تمثلان معالم التوزيع حيث أن $(\rho>٠)$ و $(٠<\alpha<١)$.
- $B_{\varepsilon}(a,b)$ تمثل دالة بيتا غير الكاملة incomplete Beta function حيث أن:

$$B_{\varepsilon}(a,b)$$
$$=\int_{٠}^{\varepsilon}x^{a-١}(١-x)^{b-١}\,.dx$$

- تنتج دالة الكثافة لتوزيع Yule بمعلمة واحدة ρ عند $(\alpha=٠)$، أي أن $\text{Yule}(\rho)=\text{GYule}(\rho,٠)$.

في دراسة [Rodríguez (٢٠١١)] تم الحصول على إمتداد آخر لتوزيع Yule وذلك من خلال خلط التوزيع الهندسي بتوزيع بيتا المعمم $[BetaI(\rho,\lambda)]$ وأطلق عليه EYule إختصارا لـ extended Yule . حيث تأخذ الدالة الإحتماليه للمتغير العشوائي Y والذي يتبع توزيع $[\text{EYule}(\rho,\lambda)]$ الشكل التالي:

$$P[Y=y]$$
$$=\frac{(\rho+١)\lambda^{y-١}B(y,\rho+١)}{{}_{٢}F_{١}(١,١;\rho+٢;\lambda)}$$
$$;\ y=١,٢,\ldots$$

حيث أن:

- ρ و λ تمثلان معالم التوزيع حيث أن $(\rho > -٢)$ و $(٠ < \lambda < ١)$.
- ${}_2F_1(\alpha,\beta;\gamma;\lambda)$ عبارة عن دالة جاوس الهندسية الزائدة، حيث أن:

$$ {}_2F_1(\alpha,\beta;\gamma;\lambda) = \sum_{r=1}^{\infty} \frac{(\alpha)_r(\beta)_r}{(\gamma)_r}\frac{\lambda^r}{r!}\ ;\ \alpha,\beta,\gamma,\lambda \in \mathbb{R} $$

- تنتج دالة الكثافة لتوزيع Yule بمعلمة واحدة ρ عند $(\lambda = ١)$ ، أي أن $\text{Yule}(\rho) = \text{EYule}(\rho, ١)$.

في هذا البحث تم الربط بين النتائج التي إنتهت إليها الدراستين [Spierdijk (٢٠٠٧)] و [Rodríguez (٢٠١١)]. ومن ثم تم الحصول على إمتداد آخر جديد لتوزيع [Yule (ρ)] بثلاثة معالم، من خلال خلط التوزيع الهندسي بتوزيع بيتا المعمم المقطوع $[BetaI(\rho,\lambda|\alpha)]$ حيث يتضمن التوزيع الجديد التوزيعات Yule و GYule و EYule كحالات خاصة مما يضيف للتوزيع ميزة هامة وهي الحصول على دقة أعلى في توفيق البيانات. كما تم أيضا الحصول على تفسير لتباين التوزيع من خلال تقسيمة إلى ثلاثة أجزاء بحسب مصادر الإختلاف (عشوائية – داخليه – خارجية).

تم إستخدام التوزيع الجديد في توفيق نفس البيانات المستخدمة في دراستي [Spierdijk (٢٠٠٧)] و [Rodríguez (٢٠١١)] و ذلك بغرض مقارنة نتائج البحث مع النتائج التي تم الحصول عليها في تلك الدراسات.

الإمتداد الجديد المقترح لتوزيع Yule

في دراسة (المقالة الأولى) تم إشتقاق دالة كثافة توزيع GYule من خلال خلط التوزيع الهندسي مع بيتا المعمم المقطوع بوضع قيد على معلمة التوزيع الثانية $(\lambda = ١)$ ، وقد تمت الإشارة لهذا الخليط بالرمز التالي:

$$ Geometric(P) \bigwedge_{P} BetaI(\rho, ١|\alpha) $$

في هذا البحث تم تقديم إمتداد آخرجديد لتوزيع Yule وذلك من خلال خلط التوزيع الهندسي مع بيتا المعمم المقطوع وبدون وضع أية قيود على المعالم، أي بإستخدام الخليط:

$$ Geometric(P) \bigwedge_{P} BetaI(\rho, \lambda|\alpha) $$

(١)

لينتج إمتداد جديد لتوزيع Yule بثلاثة معالم. يتضمن التوزيع الجديد أهم الإمتدادات المقدمه في الدراسات الحديثه [GYule, EYule] كحالات خاصة منه. حيث تم في البداية الحصول على دالة الكثافة لتوزيع بيتا المعمم المقطوع من الأسفل عند النقطة $\alpha + \lambda - ١$ بالشكل التالي:

$$f(P|\alpha) = \frac{\frac{١}{\lambda P}\left(١-\frac{١-P}{\lambda}\right)^{\rho}}{\left[\left(\frac{\lambda-\alpha}{\lambda}\right) F_{١}\left(١;-\rho;١;٢;\frac{\lambda-\alpha}{\lambda},\lambda-\alpha\right)\right]}$$

$$, \quad ١-\lambda+\alpha < P < ١$$

(٢)

حيث أن:

- (λ,ρ) هي معالم توزيع بيتا المعمم بحيث أن $(٠ < \lambda \leq ١)$ و $(\rho > -١)$.
- α هي معلمة التوزيع الشرطية حيث أن $(١-\lambda > \alpha)$.
- كما أن $F_{١}(a;b_{١},b_{٢};c;x,y)$ عبارة عن دالة Appellالهندسية

الزائدة لمتغيرين من النوع الأول، حيث أن:

$$F_{١}(a;b_{١},b_{٢};c;x,y) = \sum_{m,n=٠}^{\infty} \frac{(a)_{m+n}(b_{١})_{m}(b_{٢})_{n}}{(c)_{m+n}} \frac{x^{m}}{m!}\frac{y^{n}}{n!}$$

$$; \ |x| < ١ , |y| < ١$$

تم الحصول على دالة كثافة التوزيع الجديد من خلال إجراء التكامل التالي:

$$P(X=x) = \int_{١-\lambda+\alpha}^{١} P(١-P)^{x}. \frac{\frac{١}{\lambda P}\left(١-\frac{١-P}{\lambda}\right)^{\rho}}{\left[\left(\frac{\lambda-\alpha}{\lambda}\right) F_{١}\left(١;-\rho;١;٢;\frac{\lambda-\alpha}{\lambda},\lambda-\alpha\right)\right]}.dP$$

بوضع $(\lambda v = ١ - p)$ وعمل إزاحة للتوزيع بوضع $(Y = X + ١)$ نتج أن لدالة كثافة التوزيع الجديد الشكل التالي:

$$f(y) = \frac{(\lambda)^{y-١} B_{١-\frac{\alpha}{\lambda}}(y,\rho+١)}{\left[\left(\frac{\lambda-\alpha}{\lambda}\right) F_{١}\left(١;-\rho;١;٢;\frac{\lambda-\alpha}{\lambda},\lambda-\alpha\right)\right]}$$

$$, \quad y = ١,٢,٣,\ldots \qquad (٣)$$

تم التأكد من أن دالة الكثافة (٣) تحقق شروط دالة الكثافة الإحتمالية. وقد تم أستخدام الرمز $NYule(\rho,\lambda,\alpha)$

للإشارة لدالة كثافة التوزيع الجديد المقدمة في هذا البحث.

من المزايا الهامة لدالة الكثافة (٣) أنها تتضمن أهم الإمتدادات التي تم الحصول عليها لتوزيع Yule في الدراسات الحديثة فهو يتضمن توزيعات Yule و GYule و EYule كحالات خاصة، حيث وجد أن:

- عند $(\alpha = ٠)$ تنتج دالة كثافة التوزيع EYule المقدمة دراسة [Rodríguez, M. (٢٠١١)].
- عند $(\lambda = ١)$ تنتج دالة كثافة التوزيع GYule المقدمة دراسة [Spierdijk (٢٠٠٧)].
- عند $(\alpha = ٠, \lambda = ١)$ تنتج دالة كثافة التوزيع Yule المقدمة في دراسة [Simon (١٩٥٥)].

من خلال النتائج التي تم التوصل لها في هذا البحث يمكن الحصول على مجموعه من التوزيعات تكون عائلة لتوزيعات Yule عن طريق خلط التوزيع الهندسي بأشكال مختلفة من توزيع بيتا المعمم، الشكل (١) يلخص هذه النتائج.

خصائص توزيع $NYule(\rho, \lambda, \alpha)$

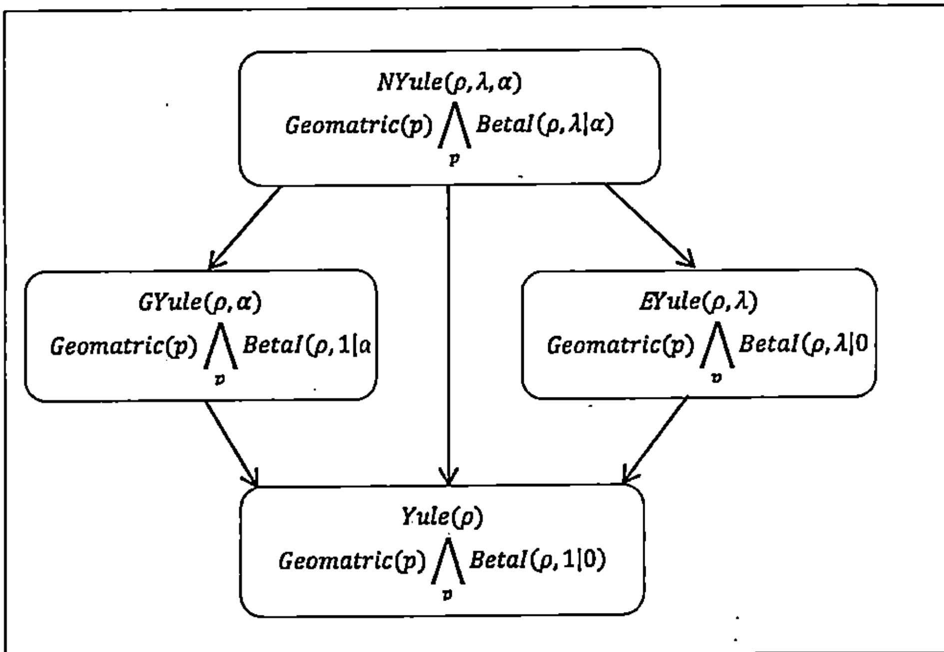
توزيع $NYule(\rho, \lambda, \alpha)$ مناسب للبيانات شديدة الإلتواء والتي يكون فيها الطرف الأيمن لمنحنى توزيعها التكراري كثيف heavy tail، حيث تكون فيها التكرارات للقيم الصغيره كبيره جدا وتنخفض بشده كلما زادت القيم لتكون التكرارات للقيم الكبيره صغيره جدا جدا. الأشكال (٢،٣،٤) توضح علاقة التوزيع بالمعالم الثلاثة (ρ, λ, α) وكيف يتأثر منحنى دالة الكثافة بها. حيث يتضح من تلك الأشكال أن حدة إلتواء التوزيع تزيد بزيادة قيم المعالم (ρ, α) وتقل بزيادة قيمة المعلمة (λ).

تم الحصول على الدالة المولدة للإحتمال لتوزيع $NYule(\rho, \lambda, \alpha)$ في هذا البحث بالشكل التالي:

$$G(z) = z\frac{F_1[١; -\rho; ١; ٢; \frac{\lambda - \alpha}{\lambda}, z(\lambda - \alpha)]}{F_1[١; -\rho; ١; ٢; \frac{\lambda - \alpha}{\lambda}, (\lambda - \alpha)]}$$

كما تم إيجاد متوسط وتباين التوزيع بالشكل التالي:

$$\mu = ١ + \frac{(\lambda - \alpha)\, F_1[٢; -\rho; ٢; ٣; \frac{\lambda - \alpha}{\lambda}, (\lambda - \alpha)]}{٢\, F_1[١; -\rho; ١; ٢; \frac{\lambda - \alpha}{\lambda}, (\lambda - \alpha)]}$$



الشكل (١): شكل توضيحي لعائلة توزيعات Yule

$$\sigma^2 = 1 + \frac{3(\lambda-\alpha)}{2}\,\frac{F_1[2;-\rho;2;3;\frac{\lambda-\alpha}{\lambda},(\lambda-\alpha)]}{F_1[1;-\rho;1;2;\frac{\lambda-\alpha}{\lambda},(\lambda-\alpha)]} + \frac{2(\lambda-\alpha)^2}{3}\,\frac{F_1[3;-\rho;3;4;\frac{\lambda-\alpha}{\lambda},(\lambda-\alpha)]}{F_1[1;-\rho;1;2;\frac{\lambda-\alpha}{\lambda},(\lambda-\alpha)]} - \mu^2$$

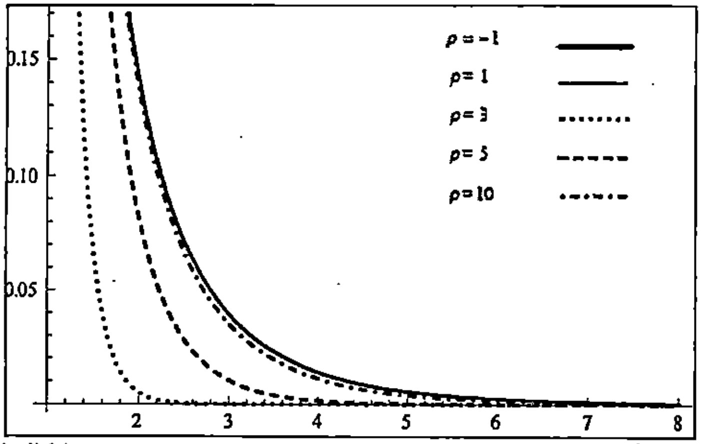
في الجزء السابق تم إشتقاق دالة الكثافة (٣) باستخدام الخليط (١)، مما يضيف ميزه أخرى للتوزيع وهي إمكانية الحصول على تفسير لتباين التوزيع من خلال تعريفه بأنه ناتج من خلط التوزيع الهندسي بدالة الكثافة:

$$P[X|p] = p(1-p)^x$$

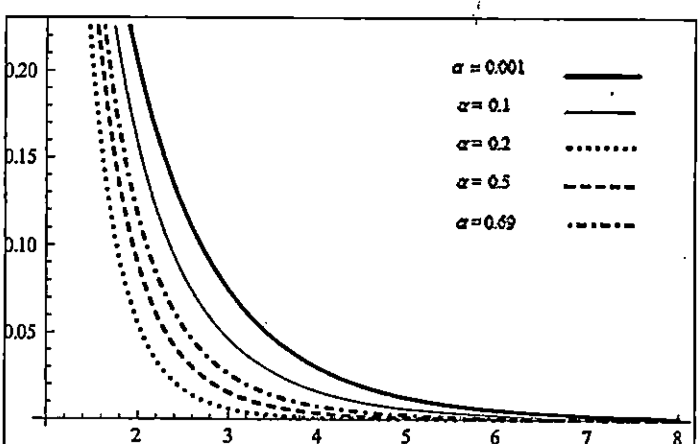
$$,x = 0,1,2,\ldots$$

$$,\quad E_{X|P}[X|p] = \frac{1-p}{p}$$

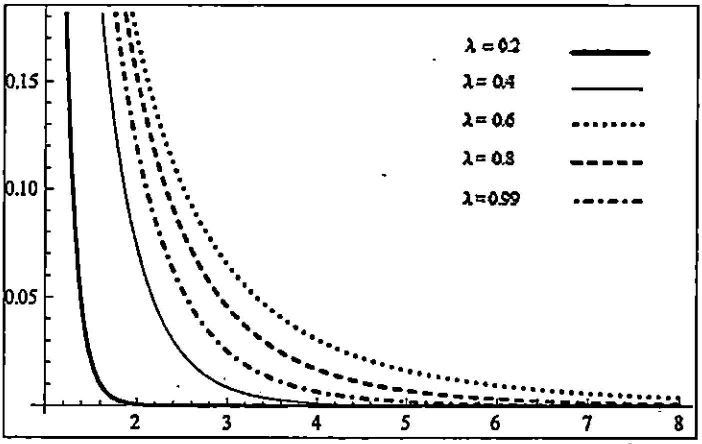
بتوزيع بيتا المعمم المقطوع من الأسفل بدالة الكثافة (٢) ، ومن ثم تم في هذا



الشكل (٢): أشكال مختلفة لدالة كثافة التوزيع $NYule(\rho, 0.7, 0.2)$ عند قيم مختلفة للمعلمة ρ



الشكل (٣): أشكال مختلفة لدالة كثافة التوزيع $NYule(2, 0.7, \alpha)$ عند قيم مختلفة للمعلمة α



الشكل (٤): أشكال مختلفة لدالة كثافة التوزيع $NYule(2,\lambda,0.2)$ عند قيم مختلفة للمعلمة λ

البحث تقسيم تباين توزيع NYule إلى ثلاثة أجزاء بحسب مصادر الإختلاف وهي كالتالي:

الأول: "randomness" أي التباين الناتج بسبب العوامل العشوائية.

الثاني: "liability" أي التباين الناتج بسبب العوامل الخارجيه.

الثالث:"proneness" أي التباين الناتج بسبب الإختلافات في المشاهدات بين المفردات.

هذه الأجزاء موضحة بالجدول (١).

تقدير معالم توزيع NewYule(ρ, λ, α)

طريقة العزوم

تم الحصول على العزوم الثلاثة الأولى للتوزيع ومساواة العزوم المحسوبة من البيانات بالعزوم النظرية ومن ثم حل المعادلات التالية:

$$a_1 = (\lambda\tau)\frac{F_1(2;-\rho;2;3;\tau,\lambda\tau)}{F_1(1;-\rho;1;2;\tau,\lambda\tau)}$$

$$a_2 = (\lambda\tau)^2\frac{F_1(3;-\rho;3;4;\tau,\lambda\tau)}{F_1(1;-\rho;1;2;\tau,\lambda\tau)}$$

$$a_3 = (\lambda\tau)^3\frac{F_1(4;-\rho;4;5;\tau,\lambda\tau)}{F_1(1;-\rho;1;2;\tau,\lambda\tau)}$$

حيث تم وضع $\left[\tau=\frac{\lambda-\alpha}{\lambda}\right]$ للتبسيط، كما أن a_i هي قيم تم حسابها من البيانات الفعلية بالشكل التالي:

$$a_1 = 2(\hat{\mu} - 1)$$

$$a_2 = \left(\frac{3}{2}\right)\left(\widehat{\mu_2} - 3\hat{\mu} + 2\right)$$

$$a_3 = \left(\frac{2}{3}\right)\Big(\left(\hat{\sigma}^3 . \widehat{s.k}\right) + 2(\hat{\mu})^3 - 3\hat{\mu}\widehat{\mu_2} - 6\widehat{\mu_2} + 11\hat{\mu} - 6\Big)$$

يمكن الحصول على مقدرات العزوم من خلال حل المعادلات السابقة عدديا .

طريقة الإمكان الأعظم

تم الحصول على دالة الإمكان للتوزيع الشكل التالي:

$$L(\theta) = n(\bar{y} - 1)\, ln[\lambda] - n\, ln[\tau] + \sum_{y=1}^{n} ln[B(y, \rho + 1) I(\tau, y, \rho + 1)] - n\, ln[\, F_1(1; -\rho; 1; 2; \tau, \lambda\tau)]$$

في هذا البحث تم الحصول على مقدرات الإمكان الأعظم عدديا وذلك بإستخدام برنامج *Mathematica*.

التطبيق العملي للإمتداد الجديد

في هذا البحث تم إعادة توفيق البيانات المستخدمة في دراستي [Spierdijk (٢٠٠٧)] و [Rodríguez (٢٠١١)] . هذه البيانات عرفت في الدراسات السابقه ببيانات النجم الساطع [superstar data]، حيث أن طبيعة هذه البيانات قائمة على فكرة التأثيرات المتراكمة [snowball effects] والتي تتسبب في وجود إلتواء شديد في التوزيع التكراري للبيانات حيث تزداد التكرارات المناظرة للقيم الصغيرة وتقل بشده كلما زادت القيم. هذا النوع من البيانات تم رصده كثيرا في مجال الفن حيث يتجه الجمهور غالبا وبتلقائية لإختيار الفنان الأكثر شهرة مما يزيد من شهرته أكثر وأكثر وذلك بغض النظر عن كونه الأفضل في مجاله [Adler (٢٠٠٦)].

البيانات المستخدمة عبارة عن ثلاثة فئات، الفئة الأولى تم توفيقها في البداية في دراسة [Chung, K. (١٩٩٤)]، هذه الفئة عبارة عن عدد النقاط الذهبيه للموسيقيين والتي تم رصدها من قبل إتحاد صناعة التسجيلات الأمريكية (RIAA) ، حيث إقتصرت الدراسة على إستخدام قائمة من البيانات المعنية بنجوم الموسيقى الشعبية أثناء الفترة (١٩٥٨-١٩٨٩) واعتبرت أن عدد النقاط الذهبيه مقياس على نجاحهم الفني. أما الفئه الثانيه والثالثة فقد تم توفيقها منذ البداية في دراسة [Giles (٢٠٠٦)] وكان مصدرها نتائج الفنانين في [Billboard Hot ١٠٠] أثناء الفترة من ١٩٥٥-٢٠٠٣. الفئة الثانية كانت عبارة عن عدد الأسابيع التي أحرزت تسجيلات الفنان فيها المركز الأول، أما الفئه الثالثه فكانت عبارة عن عدد المرات التي أحرزت تسجيلات الفنان فيها المركز الأول من بين الفنانين الذين أطلقوا ١٣ تسجيلا على الأقل أثناء تلك الفترة.

المشاهدات الفعلية الخاصة بالثلاثة فئات متاحة في كلا الدراستين السابق الإشاة لهما. النتائج التي تم الحصول عليها موضحة بالجداول (٤،٢،٣) والتي تحتوي على نتائج توفيق البيانات باستخدام توزيعات Yule باستخدام طريقة الإمكان الأعظم، بالإضافة إلى نتائج دقة التوفيق لكل توزيع. تم في هذا البحث إتباع نفس الطريقة المستخدمة في دراسة [Rodríguez (٢٠١١)] لإختبار دقة التوفيق، حيث تم استخدام إحصاءة إختبار بديلة لـ $\chi^{٢}$ من خلال عمل محاكاه لتوزيع $\chi^{٢}$ باستخدام المعاينه العشوائية من التوزيع المتقطع الخاص بالإحتمالات

التي تم توفيقها. وقد تم حساب P-value لهذه الإحصاءه باستخدام طريقة [Monte Carlo] وقد تم تنفيذ ذلك باستخدام برنامج R.

عند مستوى معنويه ٠.٠٥ ، تم رفض توزيع Yule لتوفيق بيانات الفئه الأولى و قبول التوزيعات [GYule ,EYule, NYule] لتوفيق هذه الفئة وذلك بإحتمالات متقاربة جدا هي على التوالي [٠.٢٥١,٠.٢٨٣,٠.٢٥٦]، وقد لوحظ أن تقدير المعلمة λ في توزيع NYule كان قريب جدا من الواحد وهذا يعني أن التوزيع NYule إقترب من توزيع GYule وهذا واضح من قيم المقدرات في الجدول (٢). وفي الفئه الثانية تم رفض توزيعي Yule و GYule لتوفيق البيانات وقبول توزيعي EYule و NYule لتوفيق هذه الفئة إلا أن إحتمال قبول توزيع NYule وهو [٠.٣٠٤] كان أعلى من إحتمال قبول توزيع EYule وهو [٠.٢٤٦] وذلك بالرغم من أن تقدير المعلمة α في توزيع NYule كان قريب جدا من الصفر وهذا يعني أن التوزيع NYule إقترب من توزيع EYule وهذا واضح من قيم المقدرات في الجدول (٣). أما في الفئه الثالثة فقد تم رفض توزيع Yule لتوفيق البيانات بمقدرات الإمكان الأعظم و العزوم وذلك بنفس الإحتمال تقريبا و قبول التوزيعات [GYule, EYule, NYule] لتوفيق هذه الفئة وذلك بإحتمالات متقاربة جدا هي على التوالي [٠.٤٠٧,٠.٤٣٠,٠.٤٣٩]، وقد لوحظ أن تقدير المعلمة λ في توزيع NYule كان قريب جدا من الواحد وهذا يعني أن التوزيع NYule إقترب من توزيع GYule وهذا واضح من قيم المقدرات في الجدول (٤).

من التحليل في الفقرة السابقة لنتائج توفيق البيانات في الجداول (٢،٣،٤) ، تم التأكد بالتطبيق العملي على ما تم التوصل له في هذا البحث وهو أن توزيع NYule هو الأعم و الأشمل لتوزيعات Yule المقدمة في الدراسات السابقه كما أنه يمكن الإعتماد لتوفيق البيانات شديدة الإلتواء من النوع superstar.

أيضا تم الحصول على تباين كل توزيع وتقسيمه إلى ثلاثة مكونات كما هو موضح بالجدول (٥) . حيث تم تعريف مكونات التباين لهذه البيانات بأن التباين الناتج بسبب العوامل الخارجيه يتمثل في نوع الموسيقى المقدمة أو ظروف التوسيق أو أساليب الدعاية والإعلان وغيرها من العوامل الخارجيه، أما التباين الناتج بسبب الإختلافات بين المفردات محل الدراسة تتمثل هنا في موهبة الفنان نفسه.

الجدول (٥) يحتوي على مكونات التباين وذلك لكل التوزيعات وقد لوحظ أنه لايمكن الحصول على تباين نهائي لتوزيع Yule للفئات الثلاثة والسبب هو أن مقدر معلمة التوزيع كانت أكبر من الواحد، وهذا تماما ما ذكر في دراسة [Rodríguez (٢٠١١)] إلا أنه عند تنفيذ ذلك ظهر إختلاف مهم في نتيجة حساب المكون الثاني للتباين في الفئه الثانيه للبيانات المستخدمه وهو أن نتيجة التباين كانت بالسالب [-٠.٠٤] أي أنه لا يوجد تباين نهائي لتوزيع EYule لهذه الفئه من البيانات وهذا لا يتفق مع القيمة نفسها المحسوبه في دراسة [Rodríguez (٢٠١١)] والتي كانت [٣.١٥] وذلك بالرغم من الإتفاق التام في كل القيم المحسوبه لتوزيع EYule.

وقد لوحظ أن المكون الأعلى للتباين كان بسبب العوامل الخارجية وذلك في الفئات الثلاثة، أي أن الإختلاف بين المشاهدات لم يكن عشوائيا كما أنه لم يكن ناتجا عن وجود إختلاف في قدرات الفنانين نفسهم . وهذا يتفق مع الرأي الذي تم التوصل له في دراسة [Spierdijk (٢٠٠٧), Rodríguez (٢٠١١)] والقائل بأن وجود إختلافات في مستويات النجاح بين المفردات لا يعنى بالضرورة أن هناك إختلافات بين القدرات التي تمتلكها هذه المفردات.

مناقشة النتائج

تم الحصول في هذا البحث على تعميم جديد لتوزيع Yule باستخدام أسلوب خلط التوزيعات وذلك من خلال خلط التوزيع الهندسي بتوزيع بيتا المعمم المقطوع ، حيث نتج توزيع بثلاثة معالم، مما يعطي دقه أعلى لتوفيق البيانات. بالإضافة لذلك يمكن تقسيم تباين التوزيع الجديد إلى ثلاثة مكونات (عشوائيه وخارجيه وداخلية)، بل والحصول على تباين لا نهائي للتوزيع الجديد. بعكس ما ذكر في دراسة [Rodríguez (٢٠١١)] فقد ثبت في هذا البحث أن توزيع EYule قد يعطي تباين لانهائي في بعض الحالات، حيث تم إعادة حساب تباين التوزيع لإحدى الفئات المستخدمة في الدراسة وتم التأكد من ذلك.

يمكن التوصية في هذا البحث بإمكانية الحصول على تعميم آخر لتوزيع Yule باستخدام نفس الأسلوب وذلك من خلال خلط التوزيع الهندسي بتوزيع بيتا المعمم المقطوع من الطرفين، بحيث ينتج توزيع بأربعة معالم، مما قد يعطي دقه أعلى لتوفيق البيانات.

جدول (١)

تقسيم تباين توزيع New Yule إلى العوامل الثلاثه (العشوائية والخارجية و بين المشاهدات)

Source of variability	Variance
Randomness	$\dfrac{(\lambda-\alpha)\,F_{1}\left(٢;-\rho;٢;٣;\frac{\lambda-\alpha}{\lambda},\lambda-\alpha\right)}{٢\,F_{1}\left(١;-\rho;١;٢;\frac{\lambda-\alpha}{\lambda},\lambda-\alpha\right)}$
Liability	$\dfrac{(\lambda-\alpha)^{٢}\,F_{1}\left(٣;-\rho;٣;٤;\frac{\lambda-\alpha}{\lambda},\lambda-\alpha\right)}{٣\,F_{1}\left(١;-\rho;١;٢;\frac{\lambda-\alpha}{\lambda},\lambda-\alpha\right)}$
Proneness	Liability – $(\text{Randomness})^{٢}$

جدول (٢)

لتكرارات المشاهدة والمتوقعة بإستخدام التوزيعات ***Yule*** و ***GYule*** و ***EYule*** و ***New Yule*** بالإضافة إلى نتائج إختبار $\chi^{٢}$ وحساب ***P-value*** بإستخدام طريقة ***Monte Carlo*** بتكرار ٢٠٠٠ مرة.

Gold	observed	Yule	GYule	EYule	NYule
١	٦٦٨	٧٣٢,٧٩	٦٥٨,٠٥	٦٥٤,٨٦	٦٥٨,٠٥
٢	٢٤٤	٢٣٣,٥٦	٢٤٦,٣١	٢٤٨,٢٣	٢٤٦,٣١
٣	١١٩	١١٢,٩٠	١٣١,٩٨	١٣٣,٣٨	١٣١,٩٨
٤	٧٨	٦٥,٩٣	٨٢,٤٩	٨٣,٢٥	٨٢,٤٩
٥	٥٥	٤٢,٩٧	٥٦,١٩	٥٦,٥٤	٥٦,١٩
٦	٤٠	٣٠,١٠	٤٠,٤٤	٤٠,٥٣	٤٠,٤٤
٧	٢٤	٢٢,١٩	٣٠,٢٣	٣٠,١٨	٣٠,٢٣
٨	٣٢	١٧,٠٠	٢٣,٢٣	٢٣,١١	٢٣,٢٣
٩	٢٤	١٣,٤٢	١٨,٢٣	١٨,٠٨	١٨,٢٣
١٠	١٤	١٠,٨٤	١٤,٥٥	١٤,٣٨	١٤,٥٥
١١	١٦	٨,٩٣	١١,٧٧	١١,٦٠	١١,٧٧
١٢	١٣	٧,٤٨	٩,٦٣	٩,٤٧	٩,٦٣
١٣	١١	٦,٣٥	٧,٩٥	٧,٨٠	٧,٩٥
١٤	٥	٥,٤٥	٦,٦١	٦,٤٨	٦,٦١
١٥	٤	٤,٧٣	٥,٥٣	٥,٤٣	٥,٥٣
١٦	٤	٤,١٤	٤,٦٦	٤,٥٧	٤,٦٦
١٧	٢	٣,٦٥	٣,٩٥	٣,٨٧	٣,٩٥
١٨	٧	٣,٢٤	٣,٣٦	٣,٢٩	٣,٣٦
١٩	٢	٢,٩٠	٢,٨٦	٢,٨١	٢,٨٦
٢٠	٣	٢,٦١	٢,٤٥	٢,٤١	٢,٤٥
٢١	١	٢,٣٥	٢,١١	٢,٠٨	٢,١١
٢٢	٣	٢,١٤	١,٨٢	١,٧٩	١,٨٢

٢٣	١	١٫٩٥	١٫٥٧	١٫٥٥	١٫٥٧
٢٤	١	١٫٧٨	١٫٣٦	١٫٣٥	١٫٣٦
٢٥	٠	١٫٦٤	١٫١٨	١٫١٧	١٫١٨
٢٦	٠	١٫٥١	١٫٠٣	١٫٠٢	١٫٠٣
٢٧	٠	١٫٣٩	٠٫٩٠	٠٫٩٠	٠٫٩٠
٢٨	٠	١٫٢٩	٠٫٧٨	٠٫٧٨	٠٫٧٨
٢٩	١	١٫٢٠	٠٫٦٨	٠٫٦٩	٠٫٦٨
٣٠	٠	١٫١٢	٠٫٦٠	٠٫٦٠	٠٫٦٠
٣١	٠	١٫٠٤	٠٫٥٢	٠٫٥٣	٠٫٥٢
٣٢	٠	٠٫٩٨	٠٫٤٦	٠٫٤٧	٠٫٤٦
٣٣	٠	٠٫٩١	٠٫٤٠	٠٫٤١	٠٫٤٠
٣٤	١	٠٫٨٦	٠٫٣٦	٠٫٣٧	٠٫٣٦
٣٥	٠	٠٫٨١	٠٫٣١	٠٫٣٢	٠٫٣١
٣٦	١	٠٫٧٦	٠٫٢٨	٠٫٢٩	٠٫٢٨
٣٧	١	٠٫٧٢	٠٫٢٤	٠٫٢٥	٠٫٢٤
٣٨	٠	٠٫٦٨	٠٫٢٢	٠٫٢٣	٠٫٢٢
٣٩	٠	٠٫٦٤	٠٫١٩	٠٫٢٠	٠٫١٩
٤٠	٠	٠٫٦١	٠٫١٧	٠٫١٨	٠٫١٧
٤١	٠	٠٫٥٨	٠٫١٥	٠٫١٦	٠٫١٥
٤٢	٠	٠٫٥٥	٠٫١٣	٠٫١٤	٠٫١٣
٤٣	٠	٠٫٥٢	٠٫١٢	٠٫١٣	٠٫١٢
٤٤	٠	٠٫٥٠	٠٫١٠	٠٫١١	٠٫١٠
٤٥	١	٠٫٤٨	٠٫٠٩	٠٫١٠	٠٫٠٩
$\geq$ ٤٦	١	١٨٫٨٢	٠٫٧٥	٠٫٨٩	٠٫٧٥
Estimated parameters		$\hat{\rho} =$ ١٫١٣٧	$\hat{\rho} =$ ٠٫٥٨٤ $\hat{\alpha} =$ ٠٫٠٩٠	$\hat{\rho} =$ ٠٫٤٣٤ $\hat{\lambda} =$ ٠٫٩٢٢	$\hat{\rho} =$ ٠٫٥٨٤ $\hat{\lambda} =$ ٠٫٩٩ $\hat{\alpha} =$ ٠٫٠٩٠
Goodness of fit test		$\chi^{٢} =$ ٩٣٫٢٢٤ $p =$ ٠٫٠٠١	$\chi^{٢} =$ ٥٢٫٦٥٢ p = ٠٫٢٥٦	$\chi^{٢} =$ ٥١٫٥٨٧ p = ٠٫٢٨٣	$\chi^{٢} =$ ٥٢٫٦٥٢ p = ٠٫٢٥١

جدول (٣)

التكرارات المشاهدة والمتوقعة لبيانات الفئة الثانية [Weeks] بإستخدام التوزيع GYule بإستخدام طريقة الإمكان الأعظم GYule(MLE) وطريقة العزوم GYule(MM) ، بالإضافة إلى نتائج إختبار $\chi^{٢}$ وحساب P-value بإستخدام طريقة Monte Carlo بتكرار ٢٠٠٠ مرة.

Weeks	observed	Yule	GYule	EYule	Yule
١	٣٣٧	٥١٣٫٤٨	٤١٠٫٨٧	٣٤٧٫٠٤٥	٣٥٣٫٦٩٥
٢	٢٤٩	١٦٣٫٦٧	١٩٩٫٦٠	٢٢٤٫٧١٩	٢٢٠٫٦٧٢
٣	١٣٩	٧٩٫١٢	١١٧٫٠٩	١٤٣٫٦٠٧	١٤٠٫١٥٨
٤	٩٣	٤٦٫٢٠	٧٤٫١٣	٩١٫٣٧٤	٨٩٫٥٦٥
٥	٤٧	٣٠٫١١	٤٨٫٩٦	٥٨٫٠١٣	٥٧٫٤١١
٦	٣٥	٢١٫١٠	٣٣٫٢٤	٣٦٫٧٨٥	٣٦٫٨٧٠
٧	٢١	١٥٫٥٦	٢٣٫٠٠	٢٣٫٣٠٤	٢٣٫٧٠٨
٨	١٣	١١٫٩٢	١٦٫١٥	١٤٫٧٥٥	١٥٫٢٥٨

٩	٩	٩٫٤٠	١١٫٤٦	٩٫٣٣٧	٩٫٨٢٧
١٠	٨	٧٫٦٠	٨٫٢١	٥٫٩٠٧	٦٫٣٣٢
١١	٥	٦٫٢٦	٥٫٩٣	٣٫٧٣٦	٤٫٠٨٢
١٢	٢	٥٫٢٤	٤٫٣٠	٢٫٣٦٢	٢٫٦٣٢
١٣	٢	٤٫٤٥	٣٫١٤	١٫٤٩٣	١٫٦٩٨
١٤	٤	٣٫٨٢	٢٫٣٠	٠٫٩٤٤	١٫٠٩٦
١٥	١	٣٫٣٢	١٫٧٠	٠٫٥٩٦	٠٫٧٠٧
≥١٦	٠	٤٣٫٧٣	٤٫٩٢	١٫٠٢٣	١٫٢٨٨
Estimated parameters		$\hat{\beta}$ = ١٫١٣٧	$\hat{\rho}$ = −٠٫٩٩٩ $\hat{\alpha}$ = ٠٫٢١٩	$\hat{\rho}$ = −١٫٠٢٦ $\hat{\lambda}$ = ٠٫١٣١	$\hat{\rho}$ = −٠٫٩٩٩ $\hat{\lambda}$ = ٠٫٦٤٧ $\hat{\alpha}$ = ٠٫٠٠٠
Goodness of fit test		χ^2 = ٢٦٧٫٤٨٩ p = ٠٫٠٠٠٤	χ^2 = ٤٤٫١٥٩ p = ٠٫٠٠١٥	χ^2 = ١٨٫٣٠٧ p = ٠٫٢٤٦	χ^2 = ١٧٫٢٢١ p = ٠٫٣٠٤

جدول (٤)

التكرارات المشاهدة والمتوقعة لبيانات الفئة الثالثة [Hits] بإستخدام التوزيع GYule باستخدام طريقة الإمكان الأعظم GYule(MLE) وطريقة العزوم GYule(MM) ، بالإضافة إلى نتائج إختبار χ^2 وحساب P-value بإستخدام طريقة Monte Carlo بتكرار ٢٠٠٠ مرة.

Hits	observed	Yule	GYule	EYule	NYule
١	١١٦	١٤٠٫٨٤	١٢٠٫٩٨	١٢١٫٥٦	١٢٠٫٩٨
٢	٥٧	٤٢٫٠١	٤٨٫٩٥	٤٨٫٧٣	٤٨٫٩٥
٣	٣٠	١٩٫٣١	٢٦٫٢٣	٢٦٫٠٢	٢٦٫٢٣
٤	١٣	١٠٫٨٢	١٥٫٧٦	١٥٫٦٣	١٥٫٧٦
٥	١٠	٦٫٨٢	١٠٫٠٨	١٠٫٠١	١٠٫٠٨
٦	٤	٤٫٦٣	٦٫٧١	٦٫٦٧	٦٫٧١
٧	١	٣٫٣٣	٤٫٥٩	٤٫٥٨	٤٫٥٩
٨	١	٢٫٤٩	٣٫٢١	٣٫٢١	٣٫٢١
٩	٤	١٫٩٣	٢٫٢٧	٢٫٢٨	٢٫٢٧
١٠	٢	١٫٥٣	١٫٦٣	١٫٦٤	١٫٦٣
١١	١	١٫٢٤	١٫١٨	١٫١٩	١٫١٨
١٢	٢	١٫٠٢	٠٫٨٦	٠٫٨٨	٠٫٨٦
١٣	١	٠٫٨٥	٠٫٦٤	٠٫٦٥	٠٫٦٤
١٤	١	٠٫٧٢	٠٫٤٧	٠٫٤٨	٠٫٤٧
١٥	١	٠٫٦٢	٠٫٣٥	٠٫٣٦	٠٫٣٥
١٦	١	٠٫٥٣	٠٫٢٦	٠٫٢٧	٠٫٢٦
≥١٧	٠	٦٫٣١	٠٫٨١	٠٫٨٥	٠٫٨١
Estimated parameters		$\hat{\beta}$ = ١٫٣٥٢	$\hat{\rho}$ = −٠٫٠٦٦٧ $\hat{\alpha}$ = ٠٫٢٠٤	$\hat{\rho}$ = −٠٫٠٠٥ $\hat{\lambda}$ = ٠٫٧٩٩	$\hat{\rho}$ = −٠٫٠٦٧ $\hat{\lambda}$ = ٠٫٩٩ $\hat{\alpha}$ = ٠٫٢٠٣

Goodness of fit test	$\chi^{٢} = ٣٠,٦٤٨١$ $p = ٠,٠٢٥٩$	$\chi^{٢} = ١٥,٨٤١٨$ $p = ٠,٤٣٩$	$\chi^{٢} = ١٥,٦٨٠$ $p = ٠,٤٣٠٨$	$\chi^{٢} = ١٥,٨٠٢$ $p = ٠,٤٠٧٣$

جدول (٥)

المتوسط و مكونات التباين الثلاثة بإستخدام توزيع Yule و GYule والذي تم توفيقهم بإستخدام طريقة الإمكان الأعظم وذلك للفئات الثلاثة للبيانات.

Data set	Dist	Mean	Randomness	Liability	Proneness	Total Variance
Gold records	Yule(١,١٣٨)	٨,٢٧	-	-	-	-
	GYule(٠,٥٨٤,٠,٠٩)	٣,٢٠	٢,٢٠	١٠,٥٢	٥,٦٨	١٨,٤٠
	EYule(٠,٤٣٤,٠,٩٢٢)	٣,٢٠	٢,٢٠	١٠,٦٩	٥,٨٤	١٨,٧٣
	NYule(٠,٥٨٤,٠,٩٩٩,٠,٩٠)	٣,٢٠	٢,٢٠	١٠,٥٢	٥,٦٨	١٨,٤٠
Weeks	Yule(١,١٣٧)	٢,٤٩	-	-	-	-
	GYule(٠,٠٠,٠,١٧)					
	EYule(−١,٠٢٦,٠,٦٣١)	٢,٧٥	١,٧٥	٣,٠٣	[redacted]	-
	NYule(−٠,٩٩٩,٠,٦٤٧,٠,٠٠)	٢,٧٧	٤,٧٧	٣,١٨	٠,٠٦	٥,٠٠
Hits	Yule(١,٣٥٢)	٣,٨٤	-	-	-	-
	GYule(٠,٠٠,٠,٢٠)					
	EYule(−٠,٠٠٥,٠,٧٩٩)	٢,٤٩	١,٤٩	٣,٥	١,٢٨	٦,٢٧
	NYule(−٠,٠٦٧,٠,٩٩٩,٠,٢٠٣)	٢,٤٩	١,٤٩	٣,٤٥	١,٢٣	٦,١٨

REFERENCES

١. Chung, K.H., Cox, R.A.K., (١٩٩٤), "*A stochastic model of superstardom: an application of the Yule distribution*". Review of Economics and Statistics ٧٦, p: ٧٧١-٧٧٥.

٢. Hope, A.C.A., (١٩٦٨), "*A simplified Monte Carlo significance test procedure*". Journal of the Royal Statistical Society [B] ٣٠, pp: ٥٨٢-٥٩٨.

٣. Martínez-Rodríguez, A.M., Sáez-Castillo, A.J. and Conde-Sánchez, A., (٢٠١١), "*Modelling using an extended Yule distribution*", Computational Statistics and Data Analysis ٥٥, pp: ٨٦٣-٨٧٣.

٤. Simon, H.A., (١٩٥٥), "*On a class of skew distribution functions*", Biometrika ٤٢, p: ٤٢٥-٤٤٠.

٥. Spierdijk, L. and Voorneveld, M., (٢٠٠٧), "*Superstars without talent? The Yule distribution controversy*", Working Paper Series in Economics and Finance ٦٥٨. Stockholm School of Economics.

٦. Yule, G.U., (١٩٢٥) "*A mathematical theory of evolution, based on* the conclusions of Dr. J. C. Willis", F.R.S. Phil. Trans. Roy. Soc. Lond. B ٢١٣, pp: ٢١–٨٧.

٧. conclusions of Dr. J. C. Willis", F.R.S. Phil. Trans. Roy. Soc. Lond. B ٢١٣, pp: ٢١–٨٧.