تقدير دالتي البقاء و المخاطرة لأنموذج Inverse Gompertz مع تطبيق لجانب المحاكاة

مدرس مساعد منال محمود رشيد
مديرية تربية بغداد الرصافة الثالثة
manal4254@gmail.com

تقدير دالتي البقاء و المخاطرة لأنموذج Inverse Gompertz مع تطبيق لجانب المحاكاة

مدرس مساعد منال محمود رشيد

مديرية تربية بغداد الرصافة الثالثة

manal4254@gmail.com

المستخلص :

ان الاهتمام بدراسة التوزيعات الاحتمالية من قبل الباحثين الاحصائيين جاء نتيجة للدور الذي تؤديه هذه الوسيله الاحصائية في توصيف سلوك البيانات ومعرفة خصائصها , اذ ان الظواهر يمكن التعبير عنها بمتغيرات عشوائية وكل متغير عشوائي يعبر عنه بتوزيع احتمالي يحتوي على المعلومات المهمة لهذا المتغير العشوائي .

أن التوزيـع Inverse Gompertz تم استعماله في العديد من التوزيعات الكلاسيكية على نطاق واسع على مدى العقود الماضية لنمذجة البيانات أحادية المتغير وثنائية المتغير في مجالات مثل الهندسة والعلوم الاكتوارية والبيئية والطبية والدراسات البيولوجية والديموغرافيا والاقتصاد والمالية والتأمين.

في هذا البحث تم تقدير دالتي البقاء و المخاطرة بأعتماد على طريقةالامكان الاعظم وطريقة كابلن مير (Kaplan Meier) ولأثبات امكانية تطبيق هذا الانموذج واي طريقتين افضل تم توليد حجوم عينات (n=5.10,30,80,120) وبأستعمال معيار المقارنة وهو متوسط مربعات الخطأ التكاملي (IMSE) .

الكلمات المفتاحية :

أنموذج Inverse Gompertz, دالة البقاء , دالة المخاطرة , , طريقة الامكان الاعظم , طريقة Kaplan Meier , IMSE .

1- المقدمة

تعد التجارب اختبارات الحياة من أهم الطرائق المستعملة للحصول على المعلومات المطلوبة عن الظواهر المختلفة التي تخضع لتلك التجارب. في مثل هذه التجارب ، يتم اختبار عينة من الوحدات ، ويسجل مصمم التجربة أوقات فشل وحدات العينة لأستعمالها في عمل بعض الاستنتاجات الإحصائية المفيدة في تحديد خصائص المجتمع. الوحدات المختارة من وحدات العينة وكما هو معروف أن هناك صعوبة في الحصول على أوقات الفشل للعينة بأكملها لأسباب عديدة منها التكلفة العالية أو طول فترة إجراء التجربة أو الحاجة إلى بعض وحدات العينة لتكون المستعملة في تجربة أخرى وغيرها من الاسباب.

ان الاهتمام بدراسة التوزيعات الاحتمالية من قبل الباحثين الاحصائيين جاء نتيجة للدور الذي تؤديه هذه الوسيله الاحصائية في توصيف سلوك البيانات ومعرفة خصائصها , اذ ان الظواهر يمكن التعبير عنها بمتغيرات عشوائية وكل متغير عشوائي يعبر عنه بتوزيع احتمالي يحتوي على المعلومات المهمة لهذا المتغير العشوائي .

أن التوزيع Inverse Gompertz تم استعماله في العديد من التوزيعات الكلاسيكية على نطاق واسع على مدى العقود الماضية لنمذجة البيانات أحادية المتغير وثنائية المتغير في مجالات مثل الهندسة والعلوم الاكتوارية والبيئية والطبية والدراسات البيولوجية والديموغرافيا والاقتصاد والمالية والتأمين.

من جانب اخر لكل من المعولية والبقاء خاصية واحدة ، وهي مقياس العمر الافتراضي لآلة أو نظام أو كائن حي معين ، لكن الاختلافات التي تحكمها تكمن في تحسين نظام المعولية في الأنظمة متعددة الأجزاء لأن مثل هذا التحسين ينعكس في عدد وأماكن وأجزاء هذا النظام وسهولة إيجاد بديل لهذه الأجزاء ومعالجتها. ولكن هذا التحسين غير موجود في نظرية البقاء ، لأن النظام كائن حي تكمن فيه الصعوبة والندرة في ترتيب أجزائه وإيصاله إلى حالة التحسين.

2- أنموذج Inverse Gompertz :

يعتمد هذا أنموذج على وجود معلمة واحدة لأوقات الحياة ، لنفترض بأن المتغير العشوائي X يتبع توزيع Inverse Gompertz بمعلمة مقياس واحدة $\beta(X \sim IG(\beta)$ ، فأن دالة التوزيع الاحتمالية (PDF) معطاة بالشكل الاتي :

$$f_X(x;\beta) = \frac{1}{x^2}\exp\left(\frac{1}{\beta}\left(1-\exp\left(\frac{\beta}{x}\right)\right)+\frac{\beta}{x}\right), x > 0 \qquad (1)$$

ودالة التوزيع التراكمية :

$$F_X(x;\beta) = \exp\left(\frac{1}{\beta}\left(1-\exp\left(\frac{\beta}{x}\right)\right)\right) \quad , x > 0, \beta > 0 \qquad (2)$$

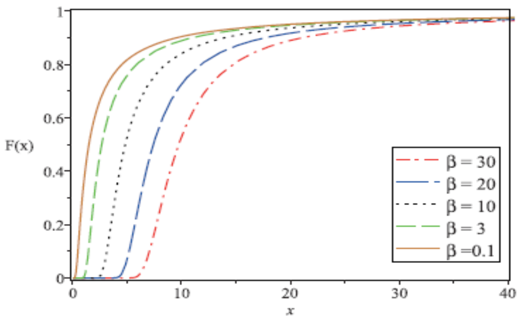

Figure 1. The CDF of A distribution for different values of β.

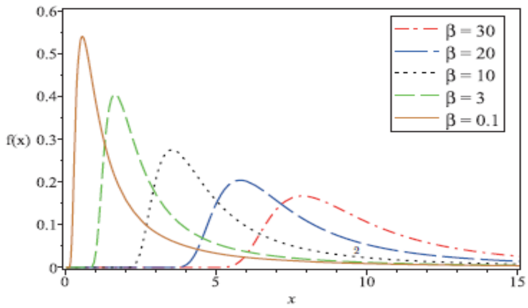

Figure 2. The PDF of A distribution for different values of β.

3- دالة البقاء :

هناك عدة معانٍ لمفهوم دالة البقاء والمعنى الواسع له هو احتمال بقاء المريض على قيد الحياة خلال مدة زمنية معينة تحت ظروف وعوامل خاصة، ويرمز لدالة البقاء بالرمز S(x)، ويعبر عن دالة البقاء بالمعادلة الرياضية الآتية :

$$S(x) = \Pr(X > x)$$

$$= \int_x^{\infty} f(y)dy$$

$$= 1 - F(x)$$

$$S_X(x;\beta) = 1 - \exp\left(\frac{1}{\beta}\left(1 - \exp\left(\frac{\beta}{x}\right)\right)\right) \quad , x > 0, \beta > 0 \tag{3}$$

4- دالة المخاطرة

دالة الخطر ، والمعروفة أيضًا باسم معدل الفشل ، هي احتمال وفاة الفرد في الوقت x نظرًا لأن الفرد نجا حتى ذلك الوقت x.
تستخدم وظيفة الخطر على نطاق واسع للتعبير عن خطر وقوع حدث (على سبيل المثال ، الموت) يحدث في وقت ما.
بالنظر إلى المتغير العشوائي X من التوزيع المستمر ، يتم إعطاء معدل الخطر h (x) بواسطة

$$h(x) = \frac{f(x)}{1 - F(x)}$$

$$h_X(x;\beta) = \frac{\exp\left(\frac{\beta}{x}\right)}{x^2 \exp\left(\frac{1}{\beta}\left(\exp\left(\frac{\beta}{x}\right) - 1\right)\right)}, x > 0 \tag{4}$$

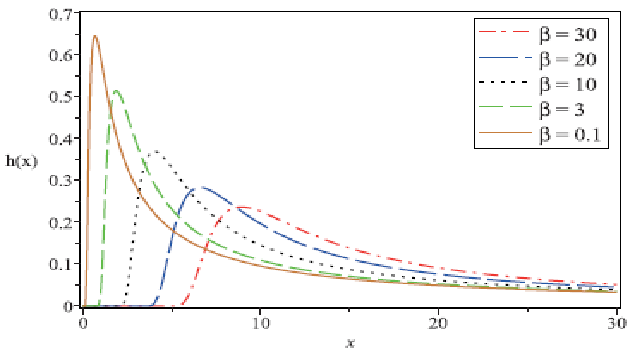

Figure 3. The HR function of the A distribution for different values of β.

5-طرائق التقدير :

1-5 طريقة الامكان الاعظم :

تعتبر هذه الطريقة من أهم الطرائق التقدير المهمة لما تحتويه من خواص جيدة ومنها الثبات ،كفاءة عالية واتساق في بعض الأحيان . افترض أن لدينا عددًا من المشاهدات من التوزيع بحجم عينة n من توزيع Inverse Gompertz. بعد ذلك ، نحصل على دالة الامكان بالشكل الاتي :

لنفترض $x_1, x_2, \dots, x_n$ تمثل عينة عشوائية بحجم n من التوزيع Inverse Gompertz فأن دالة لتوزيع الاحتمالي هي

$$\ell = \prod_{i=1}^{n} f_X(x_i;\beta)$$

$$\ell = \prod_{i=1}^{n} \frac{1}{x_i^2} \exp\left(\frac{1}{\beta}\left(1 - \exp\left(\frac{\beta}{x_i}\right)\right) + \frac{\beta}{x_i}\right)$$

$$LL = \beta \sum_{i=1}^{n} \frac{1}{x_i} - \frac{1}{\beta} \sum_{i=1}^{n} \left(\exp\left(\frac{\beta}{x_i}\right) - 1\right) - 2 \sum_{i=1}^{n} \ln(x_i)$$

$$\frac{\partial LL}{\partial \hat{\beta}} = \sum_{i=1}^{n} \frac{1}{x_i} - \frac{1}{(\hat{\beta})^2} \sum_{i=1}^{n} \left(\frac{\hat{\beta}}{x_i} \exp\left(\frac{\hat{\beta}}{x_i}\right) - \exp\left(\frac{\hat{\beta}}{x_i}\right) + 1\right) = 0$$

2-5 طريقة كابلن - مير *Kaplan Meier method*:

وهي من الطرائق اللا معلمية التي اتبعها كابلن مير (1958) *Kaplan and Meier* لتقدير دالة البقاء فهناك طريقتان الأولى يتم فيها تقدير دالة الخطورة التجميعية $H(x)$ عن طريق تقدير دالة الخطورة $h(x)$ الطريقة الثانية إستخدام دالة الكثافة التجميعية أو دالة توزيع الفشل لتقدير دالة البقاء ودالة الخطورة.

1- أسلوب كابلن - مير الأول

- ترتيب أوقات الفشل (وقت البقاء) تصاعدياً وإعطاء كل فشل رتبة (*Rank*).
- حساب دالة الخطورة $h(x)$ ، حيث أن دالة الخطورة هي عبارة عن عدد ايام اوقات البقاء مقسوم على مقدار الفترة الزمنية $=\frac{k}{n-i+1}$
$h(x)$, اذ أن k : تمثل عدد ايام اوقات البقاء , i : تمثل الرتبة , n : حجم العينة
- حساب $H(x)$ اذ أن $H(x)=[h(x_1)+h(x_2)+\ldots+h(x_n)]$
- حساب دالة البقاء $S_i(x)$, اذ ان $i=1,2$, $s_i(x) = \exp(-H_i(x))$

2- أسلوب كابلن - مير الثاني

- *تقدير* $F(x)$ التي يعتمد تقديرها على البيانات اذ ان : $F(x) = \frac{i}{n}$ او $F(x) = \frac{i-0.5}{n}$ وهي ماتسمى بدالة الكثافة التراكمية
أو $F(x) = \frac{i}{n+1}$ وهي ماتسمى برتبة الوسط الحسابي او $F(x) = \frac{i-0.3}{n+0.4}$ وهي ماتسمى برتبة الوسيط
- يمكن حساب دالة الخطورة وان دالة الفشل التجمعية حيث ان : $h_i(x) = \frac{F_{(x+1)} - F_x}{1-F_x}$
- ويمكن ايضا حساب دالة البقاء التي تساوي $S_i(x) = 1 - F_i(x)$
- ومنها نحصل على $\hat{s}(x)$ بطريقة كابلن – مير $\hat{s}_{i+1} = \prod_{i=1}^{n} s_{i+1}(x) \cdot \hat{s}_i$ اذ ان : $\hat{s}_i$ قيمة دالة البقاء المقدرة الاولى لكابلن مير
$\hat{s}_{i+1}$ قيمة دالة البقاء المقدرة اللاحقة لكابلن مير

6- الجانب التجريبي :

تم استعمال اسلوب المحاكاة بطريقة مونت كارلو للمقارنة بين طرائق التقدير المختلفة حيث يتميز هذا الاسلوب بالمرونة ويوفر الكثير من التكليف عن طريق اخذ بنظر الاعتبار حجوم العينات المختلفة والقيم المختلفة لمعلمات التوزيع وتكرار التجربة في كل مرة ويتم في هذا الاسلوب توليد البيانات دون اللجوء الى البيانات الحقيقية مع عدم الاخلال بالدقة المطلوبة وتتلخص هذه الطريقة بالخطوات التالية :

1- تحديد القيم الافتراضية :حيث تم اختيار حجوم عينات مختلفة وهي (5,10,30,80,120) وكذلك تم استعمال قيم افتراضية لمعلمة النموذج وكما يأتي :

توليد البيانات:

1- اذ تم توليد المتغير العشوائي بطريقة التحويل العكسي كمايلي :

$$x_q = \frac{\beta}{\ln(1-\beta \ln q)}$$

2- ترتيب البيانات التي تم توليدها ترتيباً تصاعدياً $x_1 < x_2 < \cdots < x_i$

3- حل المعادلات التي تم الوصول اليها بالطرق العددية

4- تم تحديد الطريقة الافضل عن طريق مقياس المقارنة $(IMSE)$ في حالة تقدير دالة البقاء والمخاطرة

$$IMSE(\hat{S}(x)) = \frac{1}{r}\sum_{i=1}^{r}\left[\frac{1}{n_x}\sum_{j=1}^{n_t}\left(\hat{s}_i(x_j) - s(x_j)\right)^2\right]$$

$$IMSE(\hat{h}(x)) = \frac{1}{r}\sum_{i=1}^{r}\left[\frac{1}{n_x}\sum_{j=1}^{n_t}\left(\hat{h}_i(x_j) - h(x_j)\right)^2\right]$$

اذ أن :

r : عدد تكرارات التجربة (1000) مرة

n_x : عدد البيانات المولدة لكل عينة

$\hat{s}_i(x_j), \hat{h}_i(x_j)$: مقدر دالة البقاء والمخاطرة على التوالي .

$S(x_j), h(x_j)$: دالة البقاء والمخاطرة حسب القيم الابتدائية وعلى التوالي .

الجدول رقم (1) يبين قيم IMSE لتقدير دالة البقاء بجميع الطرائق وحجوم العينات ولجميع التجارب

model	n	MLE	KM	best
β=2.0	5	0.025349	2.44E-02	KM
	10	8.91E-03	8.79E-03	KM
	30	8.31E-03	8.29E-03	KM
	80	8.25E-03	8.20E-03	KM
	120	1.16E-03	1.19E-03	MLE
β=2.5	5	6.76E-02	6.20E-02	KM
	10	1.19E-02	1.13E-02	KM
	30	1.17E-02	0.01141	KM
	80	1.14E-02	1.01E-02	KM
	120	8.77E-03	8.80E-03	MLE
β=4.6	5	9 .81E-03	6.82E-03	KM
	10	7.30E-03	7.06E-03	KM
	30	7.21E-03	7.00E-03	KM

	80	5.80E-03	6.81E-03	MLE
	120	1.04E-03	1.09E-03	MLE
β=5.3	5	6.21E-03	6.18E-03	KM
	10	2.45E-03	2.40E-03	KM
	30	2.44E-03	2.33E-03	KM
	80	1.61E-03	1.77E-03	MLE
	120	1.45E-05	1.51E-04	MLE

من خلال النتائج في جدول (1) تبين مايلي :

1.عند متوسط مربعات الخطأ التكاملي ولحجوم العينات المختلفة n=(5,10,30) يلاحظ وبصورة عامة إنفردت طريقة (كابلن مير) بالأفضلية على سائر طريقة الامكان الاعظم , وفي حجم عينة (n=120) ظهرت طريقة الامكان الاعظم(MLE) بالأفضلية في تقدير دالة البقاء .

2 . عند حجم عينة n=80 لنموذج الاول والثاني (β=2.0,2.5) ظهرت طريقة (كابلن مير) بالافضلية وعند نموذج الثالث والرابع (β=4.6,*5.3*) ظهرت طريقة الامكان الاعظم (MLE) بالأفضلية في تقدير دالة البقاء

الجدول رقم (2) يبين قيم IMSE لتقدير دالة المخاطرة بجميع الطرائق وحجوم العينات ولجميع التجارب

model	n	MLE	KM	best
β=2.0	5	4.29E-02	4.27E-02	KM
	10	4.22E-02	4.13E-02	KM
	30	0.041068	4.12E-02	MLE
	80	6.94E-04	4.95E-03	MLE
	120	5.22E-05	5.23E-04	MLE
β=2.5	5	6.65E-02	5.64E-02	KM
	10	6.52E-02	0.06402	KM
	30	6.42E-02	6.40E-02	KM
	80	0.0068594	6.86E-03	MLE
	120	7.53E-05	7.83E-04	MLE
β=4.6	5	0.380529	0.320761	KM
	10	0.279779	0.249545	KM
	30	0.233695	0.243841	MLE
	80	0.20355	0.233485	MLE
	120	0.020035	0.228505	MLE
β=5.3	5	0.297891	0.268753	KM
	10	0.268802	0.258384	KM
	30	0.225456	0.245317	MLE
	80	0.204796	0.215348	MLE
	120	0.0239918	0.029049	MLE

من خلال النتائج في جدول (2) تبين مايلي :

1.عند متوسط مربعات الخطأ التكاملي ولحجوم العينات المختلفة n=(5,10) يلاحظ وبصورة عامة إنفردت طريقة (كابلن مير) بالأفضلية على سائر طريقة الامكان الاعظم , وفي حجم عينة (n=80,120) ظهرت طريقة الامكان الاعظم(MLE) بالأفضلية في تقدير دالة المخاطرة .
2.عند حجم عينة n=30 لنموذج الاول والثالث والرابع (β=2.0,4.6,5.3) ظهرت طريقة الامكان الاعظم (MLE) بالافضلية وعند نموذج الثاني (β=*2.5*) إنفردت طريقة (كابلن مير) بالأفضلية في تقدير دالة المخاطرة .

7- الاستنتاجات (Conclusions) :

من خلال النتائج جانب المحاكاة تبين مايلي :
1.عند متوسط مربعات الخطأ التكاملي ولحجوم العينات المختلفة n=(5,10,30) يلاحظ وبصورة عامة إنفردت طريقة (كابلن مير) بالأفضلية على سائر طريقة الامكان الاعظم , وفي حجم عينة (n=120) ظهرت طريقة الامكان الاعظم(MLE) بالأفضلية في تقدير دالة البقاء .
2.عند حجم عينة n=80 لنموذج الاول والثاني (β=2.0,2.5) ظهرت طريقة (كابلن مير) بالافضلية وعند نموذج الثالث والرابع (β=4.6,*5.3*) ظهرت طريقة الامكان الاعظم (MLE) بالأفضلية في تقدير دالة البقاء
3.عند متوسط مربعات الخطأ التكاملي ولحجوم العينات المختلفة n=(5,10) يلاحظ وبصورة عامة إنفردت طريقة (كابلن مير) بالأفضلية على سائر طريقة الامكان الاعظم , وفي حجم عينة (n=80,120) ظهرت طريقة الامكان الاعظم(MLE) بالأفضلية في تقدير دالة المخاطرة .
4.عند حجم عينة n=30 لنموذج الاول والثالث والرابع (β=2.0,4.6,5.3) ظهرت طريقة الامكان الاعظم (MLE) بالافضلية وعند نموذج الثاني (β=*2.5*) إنفردت طريقة (كابلن مير) بالأفضلية في تقدير دالة المخاطرة .

أولاً :- المراجع العربية (Arabic References)

-1 **العامري , سماح صباح** (2021) **"تقدير دالتي البقاء والمخاطرة لتوزيع log-logistic بأستعمال الاحصاءات المرتبة مع تطبيق عملي"** رسالة ماجستير في الاحصاء مقدمة الى كلية الأدارة والأقتصاد في جامعة بغداد
-2 الباقر , زينب محمد (2017) **"تقدير دالة المعولية لتوزيع بواسون مع تطبيق عملي"** رسالة ماجستير في الاحصاء مقدمة الى كلية الأدارة والأقتصاد في جامعة كربلاء
-3 **صالح** , احمد علوان (2016) **"طرائق تقدير دالة المخاطرة لتوزيع Quasi- Lindely بحث مقارن مع تطبيق عملي"** رسالة ماجستير في الاحصاء مقدمة الى كلية الأدارة والأقتصاد في جامعة بغداد

ثانياً :- المراجع الأجنبية (Foreign References)

4- German, R.**" Non-Parametric Estimation in Survival Models"**,2005.
5- Horst, R. **"The Hazard Rate, Theory and Inference"**, With supplementary MATLAB Programs. Justus Liebig University, Germany,2014.
6- R. Alshenawy (2020) **A new one parameter distribution: properties and estimation with applications to complete and type II censored data**, Journal of Taibah University for Science, 14:1, 11-18, DOI: 10.1080/16583655.2019.1698276
7- El-Bassiouny AH, EL-Damcese M, Mustafa A, Eliwa MS. **Exponentiated generalized Weibull-Gompertz distribution with application in survival analysis**. J Stat Appl Prob. 2017;6(1):7–16.
8- El-Bassiouny AH, EL-Damcese M, Mustafa A, Eliwa MS. **Characterization of the generalized Weibull-Gompertz distribution based on the upper record values**. Int J Math Appl. 2015;3(3):13–22.
9- El-Bassiouny AH, Tahir MH, Elmorshedy M, Eliwa MS. Univariate and multivariate double slash distribution: properties and application. J Stat Appl Prob. 2019. To appear.
10-El-Bassiouny AH, EL-Damcese M, Mustafa A, Eliwa MS. Mixture of exponentiated generalized Weibull-Gompertz distribution and its applications in reliability. J Stat Appl Prob. 2016a;5(3):1–14.
11-El-Morshedy M, Eliwa MS, Nagy H. A new two-parameter exponentiated discrete Lindley distribution: properties, estimation and applications. J Appl Stat. 2019a. doi:10.1080/02664763.2019.1638893.
12-Eliwa MS, El-Morshedy M, Afify AZ. The odd Chen generator of distributions: properties and estimation methods with applications in medicine and engineering. J Natl Sci Found Sri Lanka. 2019. To appear.
13-El-Morshedy M, Eliwa MS. The odd flexible Weibull-H family of distributions: properties and estimation with applications to complete and upper record data. Filomat. 2019;33(9). To appear.
14- Basheer AM. Alpha power inverse Weibull distribution with reliability application. J Taibah Univ Sci. 2019;13(1):423–432.
15- Eliwa MS, El-Morshedy M. Discrete flexible distribution for over-dispersed data: statistical and reliability properties with estimation approaches and applications. J Appl Stat. 2020b. To appear.