تقدير الامكان الأعظم شبه المعلمي لنموذج نسبة كثافة الاحتمال لعينتين بالتطبيق علي بيانات مرضي الالتهاب الكبدي الوبائي

أ.د. محمد توفيق البلقينى
أستاذ الإحصاء الإكتوارى والتأمين
كلية التجارة- جامعة المنصورة

أ.د. فاطمة علي عبدالعاطي
أستاذ الاحصاء التطبيقي
كلية التجارة - جامعة المنصورة

لمياء علي عبدالخالق البرعي
مدرس مساعد- بقسم الإحصاء والاساليب الكمية
كلية الادارة - جامعة الدلتا للعلوم والتكنولوجيا

تقدير الامكان الأعظم شبه المعلمي لنموذج نسبة كثافة الاحتمال لعينتين بالتطبيق علي بيانات مرضي الالتهاب الكبدي الوبائي

أ.د. محمد توفيق البلقيني
أستاذ الإحصاء الإكتوارى والتأمين
كلية التجارة- جامعة المنصورة

أ.د. فاطمة علي عبدالعاطي
أستاذ الإحصاء التطبيقي
كلية التجارة- جامعة المنصورة

لمياء علي عبدالخالق البرعي
مدرس مساعد- بقسم الإحصاء والاساليب الكمية
كلية الادارة - جامعة الدلتا للعلوم والتكنولوجيا

المستخلص

تعد الطرق الحالية لتحليل البيانات المراقبة(censoring data) لنموذج نسبة الكثافة شبه المعلمي (DRM) قليلة، وذلك بسبب صعوبة الحصول علي مقدر الامكان الكامل من خلال البيانات المراقبة بسبب صعوبة الإثباتات النظرية للخصائص المتطورة، خاصة اذا كانت دالة الامكان تتطلب عدد لانهائي لمعلمات الابعاد.

يهدف هذ البحث إلى تقديم طريقة الامكان الاعظم شبه المعلمية (SML) لتقدير معلمات نموذج نسبة الكثافة لعينتين متضمنة بيانات مراقبة من اليمين (Right Censoring Data)، بالاعتماد علي بيانات تتبع التوزيع الأسي والتوزيع الطبيعي اللوغاريتمي و من خلال هذه الدراسة سوف يتم استخدام خوارزمية E-M لتقدير دالة الامكان الأعظم – لتيسير ايجاد المقدر المقترح- و يتم اجراء التقدير لكل من المكونات المعلمية و اللامعلمية للنموذج.

سوف يتم تطبيق هذا النموذج على مجموعه من البيانات المراقبة والتي تتضمن بيانات عن مرضي الالتهاب الكبدي الوبائي (فيروس C) والتي تم الحصول عليها من احدي المراكز الطبية المتخصصه. واثبتت نتائج هذه الدراسة أن النموذج المقترح يتمتع بتقديرات ذات جودة عالية وأن النموذج ملائم لطبيعة البيانات.

الكلمات المفتاحية :

نموذج نسبة الكثافة، خوارزمية E-M، تقدير الامكان الاعظم شبه المعلمي، الامكان التجريبي، البيانات المراقبة لليمين.

Abstract:

The methods currently exist for analyzing the conventional right-censored data under the semi-parametric density ratio model are considered very limited. This is mainly due to the difficulty of obtaining the full-likelihood estimator with right-censored data, the considerably complex computational algorithms and the sophisticated theoretical properties that need to be proven, especially when the likelihood function involves infinite dimensional parameters

This research aims to find semi-parametric maximum likelihood (SML) method by using the density ratio model (DRM) for two samples which include Censoring Data. Through this study the E-M algorithm will be used to estimate the maximum likelihood estimator and to facilitate finding the proposed estimator, and the estimation will be made for both parametric and nonparametric components of this model.

This model will be applied on a group of Censoring Data that include data of Hepatitis C patients obtained from a specialized medical center. the results of this study proved that the proposed model provides better fit regarding the treatment of this type of data.

Key Words: Density ratio model; EM algorithm; empirical likelihood; right-censored data; semi-parametric maximum likelihood estimation

١ . المقدمة

ظهرت في العقود الماضية الكثير من الكتابات التي تركز علي دراسة واحد من أهم النماذج شبه المعلمية وهو نموذج نسبة كثافة الاحتمال، فهو نموذج يحظي بالاهتمام لانه يحتوي علي العديد من المميزات؛ مثل: انه حل بديل لمشاكل المقارنة بين عينتين أو اكثر، حيث انه يربط توزيعين من خلال هيكل شبه المعلمي مرن لحل الافتراضات التقليدية لبعض مشاكل تعدد العينات، ويأخذ العديد من اسر التوزيعات التقليدية والنماذج كحالة خاصة.

حيث أشار Qin & Zhang (1997) الي التكافؤ بين نموذج نسبة الكثافة ونموذج الانحدار اللوجيستي لبيانات مراقبة، واقترح Qin (1998) نموذج نسبة الكثافة يتضمن نموذج العينات المتحيزة وأظهر ايضا العلاقة بين نموذج نسبة الكثافة ونموذج المخاطر النسبية Cox (1972). درس العديد من الباحثين نموذج نسبة الكثافة من جوانب عديدة مثل Zhang (2002)& Qin(2003) & Chen& Lin (2013) & Wang& Zhang (2014).

ومن هنا جاء الاهتمام بنموذج نسبة الكثافة وأهميته في ايجاد مقدراته وتطبيقة علي البيانات المراقبة .

وفى إطار ذلك تتكون هذه الدراسة من خمسه مباحث هى : المبحث الثاني يعرض الاستعراض المرجعي لنموذج نسبة الكثافة وتطبيقاتها المختلفة والمبحث الثالث تعريف نموذج نسبة الكثافة ومواصفاته. يتضمن المبحث الرابع تقدير معلمات النموذج باستخدام طريقة الامكان الاعظم. كما تم تقدير معلمات النموذج باستخدام البيانات المراقبة لليمين في المبحث الخامس، أيضا تم تطبيق التقدير المقترح لنموذج نسبة الكثافة على بيانات حقيقية لاثبات كفاءته فى المبحث السادس. واخيرا تم عرض ماتم التوصل اليه من نتائج فى هذا الدراسة فى المبحث السابع.

٢ . الدراسات السابقة:

بحث كلاً من Zhang, Chen (2021) نهجاً يعتمد على تحليل المكون الاساسي الدالي (FPCA). حيث يحدد FPCA التباين والتشابه بين الدوال ذات الأهمية من خلال التوسيع "الأمثل" لهذه الدوال. و الحصول على تقديرات تقريبية دقيقة من خلال مجموعات خطية من عدد قليل من الدوال الذاتية لنموذج DRM، وتحديد أهم الدوال الذاتية لنسب كثافة اللوغاريتم $\{(log\, g_k\,(x)/g_0(x)) : k = 0, \ldots . m\}$ من خلال عينات متعددة، وأظهرت نتائج المحاكاة التي أجريت أنه عندما يتم تعميم البيانات من التوزيعات شائعة

الاستخدام، فإن تحليل DRM على أساس دالة الأساس التكيفية له كفاءة كبيرة. في حين قدم كلا من Zhang, Zhu, Chen (2020) اختبار نسبه الامكان التجريبي للكميات لنموذج نسبة الكثافة حيث قاموا بتحديد توزيع كاي تربيع لـ ELRT لاختبار نسبه الامكان لنموذج DRM $g_k = \exp\{\theta_k^T q(x)\} g_o(x)$ ودرسوا طريقة الاستدلال للكميات عند توفر عينات متعددة من مجموعات مرتبطة. وتم اثبات ان احصائية ELRT لها توزيع كاي تربيع تحت فرض العدم في ظل ظروف معينة، ويساعد DRM على تحسين الكفاءة الإحصائية، واوضحت البيانات الحقيقية كفاءة الطريقة المقترحة. واقترح كلا من Zhuang , HU, Chen(2019) الربط بين المجتمعات المستقلة من خلال نموذج نسبة الكثافة حيث يقوم هذا النموذج بتطوير مقدر الامكان التجريبي واثبات طبيعته المتقاربة بالاضافة إلي تحسين كفاءة التقدير من خلال دراسة أوجه التشابهه بين المجتمعات وتقديم طريقة صالحة لاختبار الفرضيات وبناء فترات ثقة بناءاً علي النموذج: $d\,F_k(x) = \exp\{\theta_k^T\, q(x) dF_0(x)\}, \quad k = 1, \ldots . m$ وأظهرت نتائج المحاكاة أن المقدر المقترح يحسن بشكل كبير من كفاءة التقدير وقوة الاختبار، ويؤدي إلى فترات ثقة دقيقة بشكل كافٍ. كما أوضحت البيانات الحقيقية كفاءة الطريقة المقترحة. وتوصلت دراسة كلا من Zeng ,Gao, and Lin (2017) إلي مقدر الامكان الاعظم لنماذج الانحدار شبه المعلمي لبيانات مراقبة متعددة الفترات، واظهرت الدراسة بيانات اوقات الفشل متعدد الفترات الخاضعة للرقابة عندما تكون هناك أنواع متعددة من الفشل وكل وقت فشل يكمن في فترة معينة. حيث اعتمدت الدراسة علي الوقت الذي يؤثر في المتغيرات ربما علي اوقات الفشل متعددة المتغيرات ، من خلال النظر في فئه واسعة من نماذج التحول شبه المعلمي مع تأثيرات عشوائية، بالاعتماد علي دالة الخطر التراكمي. $\Lambda_{ijk}(t) = G_k\left[\int_0^1 exp\{\beta^T x_{ijk}^{(s)} + b_i^T z_{ijk}(s) d\,\Lambda_k(s)\}\right]$ وانتجت الدراسة ان المقدرات المقترحة للمعلمات ذات البعد المحدود متقاربة ومنتظمة بشكل طبيعي مع مصفوفة التغاير المحدود التي تحقق الكفاءة المزدوجة.

وبناء علي ما توصلت له الدراسات السابقة ووجود ندرة في تطبيق نموذج نسبة الكثافة علي البيانات المراقبة نتجت فكرة هذه الدراسة بمحاولة تقديم تقدير لنموذج نسبة الكثافة باستخدام دالة الامكان الاعظم وبالتطبيق علي البيانات المراقبة، كمحاولة من الباحثة لتقديم تقدير أكثر عمومية لتوفيق أكبر عدد ممكن من البيانات بمرونة أكثر.

٣ . نموذج نسبه الكثافة، ومواصفاته:

هو بديل شبه معلمي لحل مشكلة مقارنة عينتين أو أكثر حيث يعتمد تطبيقه بشكل كبير علي الدالة h(x) التي يفترض أنها معروفة ، حيث يؤدي التحديد الخاطئ لشكل الدالة في نموذج نسبة الكثافة إلي تقديرات متحيزة وفقدان الكفاءة. فنموذج نسبة الكثافة هو نموذج انحدار شبه المعلمي يسمح بتحليل البيانات من أي عائلة أسيه دون اجراء افتراضات التوزيع المعلمي ، حيث يتميز النموذج بالقوة في تحديد الاخطاء وكفاءة المقدر. وكما وصفة Zhang & Chen (2021) بأنه أحد الطرق الفعالة لربط التوزيعات، حيث يتمثل أحد المكونات الرئيسية في DRM في أن نسبة الكثافة اللوغاريتمية عباره عن مجموعة خطية من دوال محددة سابقاً.

بفرض أن$(X_{(1)}, X_{(2)}, \ldots, X_{(n_1)})$ عينة عشوائية مستقلة لها دالة كثافة احتمال f(x) ، وبفرض $(Y_{(1)}, Y_{(2)}, \ldots . Y_{(n_2)})$ عينة مستقلة لها دالة كثافة احتمال g(x) ولهما دالتي توزيع F(X), G(X) وكلا من (f(x), g(x)) دوال الكثافة المقابلة علي التوالي، فيكون نموذج نسبة الكثافة له الشكل التالي :

$$g(x) = exp\{(\alpha + \beta^T h(x)\} f(x) \qquad (1)$$

حيث:

α ← معلمة ثابتة غيرمعروفة للانحدار اللوجيستي

β ← متجه معاملات المتغير x في النموذج

$h(x)$ ← دالة متجه قيم المتغيرات في النموذج

$f(x)$ ← دالة كثافة الاحتمال للمتغيرات $(X_{(1)}, X_{(2)}, \ldots, X_{(n_1)})$

$g(x)$ ← دالة كثافة الاحتمال للمتغيرات $(Y_{(1)}, Y_{(2)}, \ldots . Y_{(n_2)})$

ويطلق علي نموذج نسبة الكثافة هذا الاسم لأنه يحدد النسبة اللوغاريتمية لدالتين غير معروفتين لكثافة الاحتمال الخطية في بعض المعلمات Anderson (1979) ويجدر بالذكر أن فئة التوزيعات التي ينتمي اليها هي العائلة الأسية ، ومن ثم يمكن نمذجة مجموعة كبيرة من انواع البيانات عن طريق نموذج نسبة الكثافة Konis , Fokianos (2009). لقد جذب نموذج نسبة الكثافة اهتمامًا كبيرًا الفترة الأخيرة ، لأنه يخفف العديد من الافتراضات التقليدية في سياق حل مشاكل العينات المتعددة ، حيث يتميز DRM بالمرونة ويتضمن العديد من عائلات التوزيع المعلمية ،مثل

توزيعات اللوغاريتمات وتوزيع جاما كحالات خاصة . حيث يتم التعرف علي DRM في الدراسات السابقة علي أنها أداة قوية شبه معلمية للعديد من المشاكل الاحصائية. فنموذج DRM له علاقة طبيعية في النماذج الخطية المعممة ويرتبط ارتباطاً وثيقاً بمشاكل اخذ العينات المتحيزة.

وبفرض ان لدينا متغيرين بقاء عشوائيين (X^0, Y^0) مستقلين ولهما دالتي كثافة احتمال $(g(x), f(x))$ ودالتي توزيع $G(x), F(x)$ غير معروفين، حيث نعتبر نموذج نسبة الكثافة شبه المعلمي للدالتين $g(x), f(x)$ هو :

$$g(x) = exp(\alpha + \beta^T h(x))f(x) \qquad (2)$$

عندما $\beta = 0$ تؤدي الي $\alpha = 0$ ، كذلك $\alpha = -\log \int e^{\beta^T h(x)} f(x)\, dx$ حيث $h(x)$ لها العديد من الاشكال المستخدمة في التوزيعات اكثرها استخداما : $h(x) = (x, x^2)^T$, $h(x) = log(x)$, $h(x) = x$ ومن بين هذه الاشكال ($h(x) = x$) يناسب الكثير من التوزيعات المعروفة ، مثل توزيعين أسيين لهما متوسطين مختلفين أو توزيعين طبيعين لهما نفس التباين ولكن المتوسطات مختلفة بينما ($h(x) = (x, x^2)^T$) يناسب توزيعين طبيعيين مختلفين المتوسط والتباين، ($h(x) = log(x)$) تناسب توزيعين جاما لهما نفس المعلمة او توزيعين لوغاريتم الطبيعي لهما نفس قيمة المعلمة.

في تحليل البقاء علي قيد الحياة نهتم دائماً بالاستدلال الاحصائي للبيانات المراقبة ، بفرض (X^0, Y^0) تخضعان للمراقبة العشوائية بواسطة (U, V) ، حيث (U, V) متغيرات مراقبة مستقلة لها توزيع غير معروف $H_u(x), H_v(x)$.

في هذا البحث يتم اشتقاق مقدر الامكان الاعظم شبة المعلمي (SML) لـ $\theta = (\alpha, \beta^T)^T$ ودالة التوزيع $F(x)$ لنموذج نسبة الكثافة (1). حيث تلعب خوارزمية E-M التي اقترحها Vardi (1989) دور هام جداً في تنفيذ المقدر المقترح.

٤ . طريقة الامكان الاعظم لتقدير معلمات النموذج:

يتم التقدير بطريقة الامكان التجريبي هي طريقة غير معلمية للاستدلال الإحصائي. يسمح لمحلل البيانات باستخدام طرق الامكان، دون الحاجة إلى افتراض أن البيانات تأتي من مجموعة معروفة من التوزيعات. حيث أن مزاياها في أنها تجمع بين موثوقية الطرق اللامعلمية ومرونة وفعالية نهج الامكان وتختص بتقدير المعالم من نموذج نسبة الكثافة.

بفرض ان لدينا بيانات بقاء علي قيد الحياة كالآتي: $X_i = X_i^0,\ \delta_i = 1\ ,Y_j = Y_j^0,\ \eta_j = 1$ حيث :
$i = 1,2,\dots\dots n_1\ ,\ j = 1,2,\dots\dots\dots n_2$ وبالتالي فان دالة الامكان الاعظم الكاملة هي :

$$L(\alpha,\beta,F) = \prod_{i=1}^{n_1} f(X_{1i}^0) \prod_{j=1}^{n_2} g(Y_j^0)$$

$$= \prod_{i=1}^{n_1} f(X_i^0)\ \prod_{j=1}^{n_2} f(Y_j^0)\, exp(\alpha + \beta^T h(X)) \qquad (3)$$

وبفرض أن: $b_{1k} = \sum_{i=1}^{n_1} I(X_i^0 = t_k)$, $b_{2k} = \sum_{j=1}^{n_2} I(Y_j^0 = t_k)$ وعند الحدث

$t_k (k = 1,\dots\dots\dots K)$ في مجموعتي المتغيرات، حيث تلاحظ القيم المميزة للمشاهدات بين المجموعات متزايدة
$t_1 < t_2 < \cdots \dots < t_k$ حيث يلاحظ أن القيم للوقت t_k في المشاهدات بين المجموعات تكون موجودة بترتيب
تصاعدي ، وبفرض : $b_k = b_{1k} + b_{2k}$ وبالتالي فإن: $\sum_{k=1}^{K} b_{1k} = n_1$; $\sum_{k=1}^{K} b_{2k} = n_2$ وأيضاً
$\sum_{k=1}^{K} b_k = n$ ، فان دالة الامكان الكاملة يمكن كتابتها كما يلي :

$$L(\alpha,\beta,F) = \prod_{k=1}^{K} f(t_k)^{b_{1k}}\, g(t_k)^{b_{2k}}$$

$$= \prod_{k=1}^{K} f(t_k)^{b_{1k}} \left(exp(\alpha + \beta^T h(t_k)).f(t_k)\right)^{b_{2k}}$$

$$= \prod_{k=1}^{K} f(t_k)^{b_{1k}} \left(exp(\alpha + \beta^T h(t_k))\, b_{2k}\, .f(t_k)\right)^{b_{2k}}$$

$$= \prod_{k=1}^{K} f(t_k)^{b_k}\ exp\{(\alpha + \beta^T h(t_k))b_{2k}\}$$

وتكون دالة لوغاريتم الامكان هي :

$$l(\alpha,\beta,F) = \sum_{k=1}^{K}\{b_k \log f(t_k) + b_{2k}\,(\alpha + \beta^T h(t_k))\} \qquad (4)$$

وكما ذكر Vardi (1989) الي تعظيم (4) لأنها كافية وللنظر في دالة التوزيع المنفصلة F التي لديها الكتل الايجابية تقتصر علي هذه النقاط $\{t_1, t_2, \dots\dots\dots t_k\}$ فقط وحيث أن :

$$f(t_k) = p_k \quad , k = 1, \dots\dots\dots K, \;\; p = (p_1, \dots\dots\dots\dots p_k)$$

فتكون دالة لوغاريتم الامكان بالنسبة للمعلمات (α, β, p) كما يلي :

$$l(\alpha, \beta, p) = \sum_{k=1}^{K} \{b_k \log\, p_k + b_{2k}\,(\alpha + \beta^T h(t_k))\} \qquad (5)$$

وطبقاً لطريقة كلاً من Qin& Lawless (1994) نحن نعظم (5) طبقاً للشروط :

$$\sum_{k=1}^{K} p_k = 1 \,, \quad 0 \le p_k \le 1 \quad , \; k = 1, \dots\dots\dots K$$

$$, \sum_{k=1}^{K} exp\{\alpha + \beta^T h(t_k)\}\, p_k = 1 \qquad (6)$$

مما يضمن أن كلاً من $f(x)$ ، $g(x)$ هي دوال التوزيع وباستخدام طريقة مضاعف لاجرانج Lagrange multiplier method تتحقق القيمة العظمي للقيم $l(\alpha, \beta, p)$ عند:

$$\tilde{p}_{\,k} = \frac{b_k}{n_1 + n_2\, exp\left(\alpha + \beta^T h(t_k)\right)}$$

حيث $k = 1,2, \dots\dots\dots K$ للثوابت (α, β)، وبالتالي تجاهل العناصر الثابتة ، تعرف دالة لوغاريتم الامكان لـ (α, β)هي :

$$l(\alpha, \beta) = \sum_{k=1}^{K} \{b_k \log\, p_k + b_{2k}\,(\alpha + \beta^T h(t_k))\}$$

$$= \sum_{k=1}^{K} \left\{ b_k \log \frac{b_k}{n_1 + n_2\, exp\left(\alpha + \beta^T h(t_k)\right)} + b_{2k}\,(\alpha + \beta^T h(t_k)) \right\}$$

$$= \sum_{k=1}^{K} \left\{ b_k \log b_k - b_k \log(n_1 + n_2\, exp\left(\alpha + \beta^T h(t_k)\right) + \sum_{k=1}^{K} b_{2k}(\alpha + \beta^T h(t_k)) \right\}$$

$$= -\sum_{k=1}^{K} \left\{ b_k \log(n_1 + n_2\, exp\left(\alpha + \beta^T h(t_k)\right) + \sum_{k=1}^{K} b_{2k}(\alpha + \beta^T h(t_k)) \right\}$$

وبحساب التفاضل لدالة لوغاريتم الامكان بالنسبة لكل من (α, β) والحصول علي مقدراتهم $(\tilde{\alpha}, \tilde{\beta})$ فان :

$$l(\alpha,\beta) = \sum_{k=1}^{K} b_{2k}\left(\alpha + \beta^T h(t_k)\right) - \sum_{k=1}^{K}\left\{ b_k \log\left(n_1 + n_2 \exp\left(\alpha + \beta^T h(t_k)\right)\right)\right\} \quad (7)$$

$$\frac{\partial l}{\partial \alpha} = \sum_{k=1}^{K} b_{2k} - \sum_{k=1}^{K} \frac{b_k n_2 \exp\left(\alpha + \beta^T h(t_k)\right)}{n_1 + n_2 \exp\left(\alpha + \beta^T h(t_k)\right)}$$

$$\frac{\partial l}{\partial \alpha} = n_2 - \sum_{k=1}^{K} \frac{b_k n_2 \exp\left(\alpha+\beta^T h(t_k)\right)}{n_1+n_2 \exp\left(\alpha+\beta^T h(t_k)\right)} = 0$$

$$\frac{\partial l}{\partial \beta} = \sum_{k=1}^{K} b_{2k} h(t_k) - \sum_{k=1}^{K} \frac{b_k n_2 h(t_k) \exp\left(\alpha + \beta^T h(t_k)\right)}{n_1 + n_2 \exp\left(\alpha + \beta^T h(t_k)\right)} = 0$$

ولذلك نحصل علي :

$$\tilde{p}_k = \frac{b_k}{n_1+n_2 \exp\left(\tilde{\alpha}+\tilde{\beta}^T h(t_k)\right)}$$

حيث : $k = 1,2,\ldots\ldots K$ وبالاعتماد علي مقدر الامكان الاعظم شبه المعلمي SML $(\tilde{\alpha},\tilde{\beta})$ و $\tilde{p}_k$ و $K = 1,2,\ldots\ldots k$ فأن تقدير كلاً من $G(x), F(x)$ هو :

$$\tilde{F}(x) = \sum_{k=1}^{K} \tilde{P}_k I(t_k \le x)$$

$$\tilde{G}(x) = \sum_{k=1}^{K} \tilde{P}_k \exp\left\{\left(\tilde{\alpha} + \tilde{\beta}^T h(t_k)\right)\right\} I(t_k \le x) \qquad (8)$$

٥ . تقدير المعلمات للبيانات المراقبة من اليمين: The right censored data

في هذا الجزء سنقوم بتطبيق المقدر السابق الي حالة البيانات المراقبة من اليمين ولبيانات بقاء مراقبة من اليمين $\{(X_i,\delta_i): i = 1,\ldots\ldots,n_1\}$ و $\left\{\left(Y_j,\eta_j\right): j = 1,\ldots\ldots,n_2\right\}$ وتكون دالة الامكان الكاملة هي :

$$L(\alpha,\beta,F) = \prod_{i=1}^{n_1} f(X_i)^{\delta_i} \left(1 - F(X_i)^{1-\delta_i} \cdot \prod_{j=1}^{n_2} g\left(Y_j\right)^{\eta_j} \left(1 - G(Y_j)\right)^{1-\eta_j}\right.$$

$$= \prod_{i=1}^{n_1} f_1(X_i)^{\delta_i} (1 - F_1(X_i))^{1-\delta_i} \cdot \prod_{j=1}^{n_2} f_2(Y_j)^{\eta_j} exp\left\{(\alpha + \beta^T h(Y_j))\eta_j\right\} (1 - F_2(Y_j))^{1-\eta_j}$$

وبفرض ($t_1 < t_2 < \cdots \ldots < t_k$) تدل علي قيم متميزة بما في ذلك الخاضعة للمراقبة وغير الخاضعة للمراقبة حيث $K = 1, \ldots \ldots, k$ ، وبفرض :

$$r_{1k} = \sum_{i=1}^{n_1} I(X_i = t_k, \delta_i = 1) \, , \, \xi_{1k} = \sum_{i=1}^{n_1} I(X_i = t_k, \delta_i = 0)$$

$$r_{2k} = \sum_{j=1}^{n_2} I(Y_j = t_k, \eta_j = 1) \, , \, \xi_{2k} = \sum_{j=1}^{n_2} I(Y_j = t_k, \eta_j = 0)$$

وتكون المضاعفات لكل المشاهدات المراقبة وغير المراقبة في الوقت t_k في المجموعتين علي التوالي ، ثم يمكن توضيح دالة الامكان الكاملة كما يلي:

$$L(\alpha, \beta, F) = \prod_{k=1}^{K} f_0(t_k)^{r_{1k}+r_{2k}} exp\{(\alpha + \beta^T h(t_k))r_{2k}\} (1 - F_0(t_k))^{\xi_{1k}} (1 - F_1(t_k))^{\xi_{2k}}$$

وتكون دالة الامكان اللوغاريتمية هي كما يلي :

$$l(\alpha, \beta, F) = \sum_{k=1}^{K}\{(r_{1k}+r_{2k}) \log f_0(t_k) + r_{2k} (\alpha + \beta^T h(t_k)) + \xi_{1k} log(1 - F_0(t_k)) + \xi_{2k} log(1 - F_1(t_k))\} \quad (9)$$

ويتم التركيز علي دالة التوزيع المنفصلة F التي لها كتله إيجابية محصورة علي النقاط $\{t_1, t_2, \ldots \ldots, t_k\}$ للحصول علي تعظيم (9) وبالمثل تدل علي:

$$p_k = f_0(t_k), \quad q_k = f_1(t_k) = exp\{(\alpha + \beta^T h(t_k))\} f_0(t_k)$$

$$k = 1, \ldots \ldots, K \qquad , p = (\ p_1, \ldots \ldots \ldots \ldots, p_k)$$

فان دالة لوغاريتم الامكان بالنسبة (α, β, p) هي :

$$l(\alpha, \beta, p) = \sum_{k=1}^{K}\{(r_{1k}+r_{2k}) \log p_k + r_{2k} (\alpha + \beta^T h(t_k)) + \xi_{1k} log(\sum_{l=k+1}^{k} p_l) + \xi_{2k} log(\sum_{l=k+1}^{k} exp(\alpha + \beta^T h(t_l) p_l)\} \quad (10)$$

ولتعظيم (10) تخضع للشروط:

$$\sum_{k=1}^{K} p_k = 1 \ , \ 0 \le p_k \le 1 , \ k = 1, \dots \dots , K$$

$$, \sum_{k=1}^{K} exp\{\alpha + \beta^T h(t_k)\} p_k = 1 \qquad (11)$$

ويمكننا الحصول علي مقدر SML (α, β, p) حيث نعرض مقدر SML بشكل فريد في ظل بعض الشروط المنتظمة ، حيث تعظيم (10) بطريقة مباشرة مهمه صعبة، وبالتالي نلجأ هنا إلي خوارزمية حسابية ممكنه لتنفيذه بسهولة وهي خوارزمية E-M، التي اقترحها المؤلف Vardi(1989) للحصول علي مقدر الامكان الاعظم غير المعلمي لدوال البقاء علي قيد الحياة مع بيانات مراقبة مضاعفة الآن نحن بحاجه الي حساب المتوقع من كل تكرار وفي خطوة M حيث نقوم بتحديث مقدر لهذه المعلمات ، ويمكننا ان نبدأ بالتكرار :

$\alpha^{(0)}, \beta^{(0)}$ $p^{(0)} = (p_1^{(0)}, \dots \dots . p_K^{(0)})$, التي تحقق الشروط في (11) وللتبسيط نحن دائماً نضع القيمة الاولية $\alpha^{(0)} = 0 \ , \beta^{(0)} = 0 , p_1^{(0)} = \cdots = p_K^{(0)} = \frac{1}{K}$ وخطوات mth تتم عملية التكرار وفقا للإجراءات التالية :

أولاً : خطوة التوقع E-step:

$p^{(m-1)} = (p_1^{(m-1)}, \dots \dots . p_K^{(m-1)})$, $\alpha^{(m-1)}, \beta^{(m-1)}$ مع n من المشاهدات و المعلمات المقدرة والحصول عليها من التكرار (m-1)th فان التكرار الشرطي المتوقع للمجموعة الاولي هو كالتالي:

$$b_{1k}^{(m)} = \sum_{i=1}^{n_1} E\left[I(X_i^o = t_k , \delta_i = 1) \,\middle|\, (X_i, \delta_i), \alpha^{(m-1)}, \beta^{(m-1)}, p^{(m-1)}\right]$$

$$+ \sum_{i=1}^{n_1} E\left[I(X_i^o = t_k, \delta_i = 0) \,\middle|\, (X_i, \delta_i), \alpha^{(m-1)}, \beta^{(m-1)}, p^{(m-1)}\right]$$

$$= r_{1k} + \sum_{i=1}^{n_1} E\left[I(X_i^o = t_k, \delta_i = 0) \,\middle|\, (X_i, \delta_i), \alpha^{(m-1)}, \beta^{(m-1)}, p^{(m-1)}\right]$$

حيث:

$$E\left[I(X_i^O = t_k\,, \delta_i = 0)\,\middle|\,(X_i\,, \delta_i), \alpha^{(m-1)}\,, \beta^{(m-1)}, p^{(m-1)}\right]$$

$$= I(X_i^O \le t_k\,, \delta_i = 0)\frac{P(X_i^O = t_k\,, U_i = X_i)}{\sum_{l:t_1 \ge X_i} P(X_i^O = t_l\,, U_i = X_i)}$$

$$= I(X_i \le t_k\,, \delta_i = 0)\frac{f(t_k)\, d\, H_u(X_i)}{\sum_{l:t_1 \ge X_i} f(t_l) dH_u(X_i)}$$

$$= p_K^{(m-1)} \sum_{j=1}^{k} I\big(X_i = t_j\,, \delta_i = 0\big)\frac{1}{\sum_{l:t_l \ge X_i} p_l^{(m-1)}}$$

وتأتي المعادلة الثانية من فرض المراقبة المستقلة ، نحصل علي :

$$b_{1k}^{(m)} = r_{1k} + \sum_{j=1}^{k} I\big(X_i = t_j\,, \delta_i = 0\big)\frac{1}{\sum_{l:t_l \ge X_i} p_l^{(m-1)}} + \sum_{i=1}^{n_1} p_K^{(m-1)}$$

$$= r_{1k} + p_K^{(m-1)} \sum_{j=1}^{k} \sum_{i=1}^{n_1} \frac{I\big(X_i = t_j\,, \delta_i = 0\big)}{\sum_{l=j}^{k} p_l^{(m-1)}}$$

$$= r_{1k} + p_K^{(m-1)} \sum_{j=1}^{k} \frac{\xi_{1j}}{\sum_{l=j}^{k} p_l^{(m-1)}}$$

وبالمثل فان التكرار الشرطي المتوقع للمجموعة الثانية هو :

$$b_{2k}^{(m)} = r_{2k} + q_K^{(m-1)} \sum_{j=1}^{k} \frac{\xi_{2j}}{\sum_{l=j}^{k} q_l^{(m-1)}}$$

حيث:

$$q_k^{(m-1)} = exp\big\{\alpha^{(m-1)} + \beta^{(m-1)} h(t_k)\big\} p_k^{(m-1)}$$

وفيما يلي: نقوم بتحديث المقدرات باستخدام مجموعة البيانات الكاملة الجديدة:

$$\{b_{1k}^{(m)}, b_{2k}^{(m)}, t_k; k = 1, \dots\dots, K\}$$

ثانياً : خطوة التعظيم M- step:

وفقاً للمعادلة (11) فان دالة لوغاريتم الامكان للـ $\{b_{1k}^{(m)}, b_{2k}^{(m)}, t_k; k = 1, \dots\dots, K\}$ هي كما يلي:

$$l^m(\alpha, \beta, p) = \sum_{k=1}^{K} \{(b_{1k}^{(m)} + b_{2k}^{(m)}) \log p_k + b_{2k}^{(m)}(\alpha + \beta^T h(t_k))\}$$

وباستخدام مضاعف لاجرانج Lagrange Multiplier مرة اخري يكون الحد الاقصي لدالة لوغاريتم الامكان التي تحقق الشروط (11) هي كما يلي :

$$p_K^{(m)} = \frac{b_{1k}^{(m)} + b_{2k}^{(m)}}{\omega_1^{(m)} + \omega_2^{(m)} exp(\alpha^{(m)} + \beta^{(m)T} h(t_k))}, \quad k = 1, \dots, K,$$

حيث :

$$\omega_1^{(m)} = \sum_{k=1}^{K} b_{1k}^{(m)} \quad , \quad \omega_2^{(m)} = \sum_{k=1}^{K} b_{2k}^{(m)}$$

و α^m, β^m هي جذر معادلات النتيجة التالية :

$$\frac{\partial l^m}{\partial \alpha} = \omega_2^{(m)} - \sum_{k=1}^{K} \frac{(b_{1k}^{(m)} + b_{2k}^{(m)}) \omega_2^{(m)} exp(\alpha + \beta^T h(t_k))}{\omega_1^{(m)} + \omega_2^{(m)} exp(\alpha + \beta^T h(t_k))} = 0$$

$$\frac{\partial l^m}{\partial \beta} = \sum_{k=1}^{K} b_{2k}^{(m)} h(t_k) - \sum_{k=1}^{K} \frac{(b_{1k}^{(m)} + b_{2k}^{(m)}) \omega_2^{(m)} h(t_k) exp(\alpha + \beta^T h(t_k))}{\omega_1^{(m)} + \omega_2^{(m)} exp(\alpha + \beta^T h(t_k))} = 0$$

نكرر كلا من الخطوة E والخطوة M بالتناوب حتي يتقارب كلاً من :

$$\theta^m = (\alpha^{(m)}, \beta^{(m)T})^T, p_K^{(m)}, k = 1, \dots, K$$

والتقارب يمكن تطبيق عدة معايير علي الدراسات العددية حيث سيتم انهاء التكرار عندما تكون l_2 معيار الفرق بين اثنين من مقدرات التكرارات المجاورة أقل من الحد الأقصي المحدد. ونحصل اخيراً علي مقدر SML للبيانات المراقبة ويشار اليه:

$\hat{\theta} = (\hat{\alpha}, \hat{\beta}^T)^T$, $\hat{P} = (\hat{P_1}, \dots \dots \hat{P_K})$ ، ومقدر SML لـ $F(x)$ و $G(x)$ هي:

$$\hat{F}(x) = \sum_{k=1}^{K} \hat{p}_k \, I(t_k \le x) \, ,$$

$$\hat{G}(x) = \sum_{k=1}^{K} \hat{p}_k \, exp\{\hat{\alpha}, \hat{\beta}^T h(t_k)\} I(t_k \le x), \qquad (12)$$

٦ . الجانب التطبيقي:

نتناول في هذا الجزء التطبيق العملي لنموذج شبه المعلمي DRM لعينتين بافتراض المراقبة من جهة اليمين علي بيانات واقعية تمثل عينة من مرضي الالتهاب الكبد الوبائي المزمن ((Hepatitis C virus (HCV)). تشتمل البيانات علي عدد 379 مريض وعدد من المتغيرات التي تم قياسها فتضمنت الدراسة علي عدد 379 مريضاً بالغًا يعانون من النوع الجيني 4 من فيروس الالتهاب الكبدي الوبائي المزمن(Hepatitis C) والذين تم تعيينهم لتلقي العلاج من التهاب الكبد الفيروسي مع تركيبة جرعة ثابتة من (75/50 / 12.5) OBV / PTV / RBV مجم، بمعدل قرصين مرة واحدة يوميًا علي أن تكون الجرعة علي أساس الوزن ((1000) RBV مجم / يوم) إذا كان وزن الجسم أقل من ٧٥ كجم أو جرعة (1200 مجم / يوم) إذا كان الوزن أكثرمن 75 كجم، مع أو بدون 400 مجم من SOF يوميًا (لأولئك الذين فشلوا في العلاج السابق بـ daclatasvir (DCV) plus sofosbuvir)، وذلك خلال الفترة ما بين يناير 2016 ويونيو 2018، من مركزعلاج التهاب الكبد الفيروسي بالقاهرة الجديدة مركز (NCVHTC)؛ أحد مراكز العلاج المتخصصة التابعة للجنة الوطنية لمكافحة التهاب الكبد الفيروسي (NCCVH) في مصر. تم تحديد وقت لجميع المرضى المعنيين

لتلقى العلاج RBV / PTV / OBV لمدة 12 أسبوعًا وفقًا لما إذا كانوا لم يسبق لهم أي علاج بأدوية الالتهاب الكبدي الفيروسي (Treatment Naïve) (مشاهدات غير مراقبه) أو 24 أسبوعاً (Treatment Experienced) (مشاهدات مراقبة) إذا كانوا سبق لهم العلاج بأدوية أخري للالتهاب الكبدي الفيروسي، وذلك وفقًا للمعاييرالموحدة لبروتوكول صادر عن مركز (NCCVH). (El Kassas 2019)

من هذه البيانات سوف نعتمد علي ثلاث متغيرات للتطبيق في نموذج نسبة الكثافة شبه المعلمي بين مجموعتين في حالة المراقبة جهة اليمين، وهذه المتغيرات هي:

❖ جدول (١) : وصف المتغيرات المستخدمة فى البحث

اسم المتغير بالانجليزية	اسم المتغير باللغة العربية	التعريف
Gender	النوع	ويمثل نوع المريض الذكر بالكود 1 بينما نوع المريض الأنثي بالكود 0.
Treatment	المعالجة	المعالجة الأولي : "Treatment Naïve" وهي عبارة عن المرضي الذين لم يسبق لهم تداول أي علاج للالتهاب الكبدي الوبائي الفيروسي، وسوف يتم تكويدها بالرقم 1 المعالجة الثانية: "Treatment Experienced" وهي عبارة عن المرضي الذين سبق لهم تداول علاجات أخري للالتهاب الكبدي الوبائي الفيروسي وسوف يتم تكويدها بالرقم 0.
Total BILIRUBIN	البيليروبين	وهي عبارة عن نسبة الصفراء في الدم. وهو يمثل المشاهدات (x). ونقوم بتقسيم المشاهدات الخاصة بمجموعة الذكور بـ x_1 وعددها n_1 بينما المشاهدات الخاصة بمجموعة الأناث سوف نرمز لها بالرمز x_2 وعددها n_2.

في النموذج شبه المعلمي المقترح DRM يمثل متغير النوع المجموعتين حيث أن المجموعة الأولي تمثل عينة المرضي الذكور وعددها 218 مريض بنسبة مئوية 57.5% بينما المجموعة الثانية تمثل عينة المرضي الأناث وعددها 161 بنسبة مئوية 42.5%. بينما يمثل متغير المعالجة نوع المراقبة من جهة اليمين حيث أن المعالجة Treatment Naïve تمثل الوحدات غير المراقبة وعددها 327 مشاهدة بنسبة مئوية 86.3% بينما المعالجة Treatment Experienced تمثل الوحدات المراقبة وعددها52 بنسبة مئوية 13.7%.

وبالتالي يكون لدينا جدول الأقتران للعلاقة بين نوع المعالجة المستخدمة ومتغير النوع كما يلي

Table

	Female	Male
Treatment Experienced	17	35
Treatment Naive	144	183

ومن الجدول السابق فأن نسبة المراقبة في المجموعة الأولي تمثل عدد الذكور تحت المراقبة ما يعادل تقريبا نسبة 60% بينما نسبة المراقبة في المجموعة الثانية تمثل عدد الأناث تحت المراقبة أي ما يعادل تقريبا نسبة 40%. نحدد دالة مؤشرلكل مجموعة (المجموعة الأولي المؤشر لها I_1 والمجموعة الثانية المؤشر لها I_2) وهذا المؤشر يأخذ قيمة 1 أذا كانت الوحدات مراقبة ويأخذ قيمة 0 إذا كانت الوحدة غير مراقبة.

حيث أن فروض الدراسة هي:

الفرض العدمي H_0: لا يوجد اختلاف بين المجموعتين (ذكور وإناث) تحت نوع المعالجة المستخدم

الفرض البديل H_1: يوجد اختلاف بين المجموعتين (ذكور وإناث) تحت نوع المعالجة المستخدم

فإذا كانت الأحتمالية المحسوبة للاحصائية ll أقل من 5% فأن ذلك دليل علي قبول الفرض البديل ورفض الفرض العدمي مما يعني أن هناك اختلاف معنوي ذو دلالة إحصائية بين المجموعتين (الذكور والأناث) تحت نوع المعالجة المستخدم وباعتبار المراقبة جهة اليمين. أما اذا كانت الأحتمالية المحسوبة للإحصائية ll أكبر من 5% فأن ذلك دليل علي رفض الفرضية البديلة وقبول الفرض العدمي بمعني أنه لا يوجد اختلاف بين مجموعتي (الذكور والأناث) تحت نوع المعالجة المستخدم.

في الجدول التالي نستعرض نتائج تقديرات المعالم (α,β)والخطأ المعياري وكذلك فترة الثقة لكل معلمة عند مستوي% 95 وتقدير R_n في حالة أن دالة الفشل تتبع التوزيع الأسي والتوزيع اللوغاريتم الطبيعي بالتطبيق علي البيانات الواقعية .

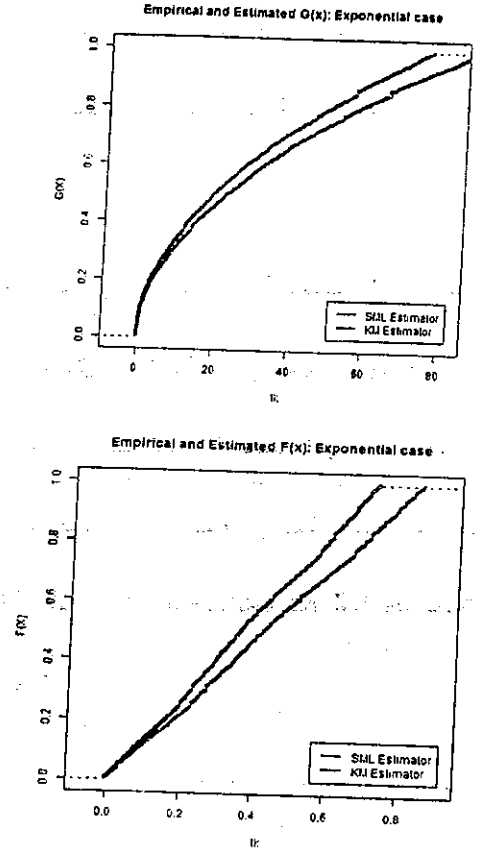
جدول رقم (٢): نتائج البيانات الحقيقية (المتوسط والخطأ المعياري وفترة الثقة عند مستوي 95%) للتقديرات شبه المعلمية للمعلمتين (α,β) وكذلك تقدير حجم نسبة الأمكان تبعاً لنوع التوزيع المفترض

التوزيع	تقدير التحيز		تقدير الخطأ المعياري		تقدير فترة الثقة		تقدير R_n
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	p-value
الأسي	0.1996	0.2271	0.4308	0.9165	(-0.2556, 0.9622)	(-0.6579, 0.7802)	0.01254
اللوغاريتم الطبيعي	0.3465	0.0533	0.3631	0.3161	(-0.2738, 0.9679)	(-0.6613, 0.7669)	0.01281

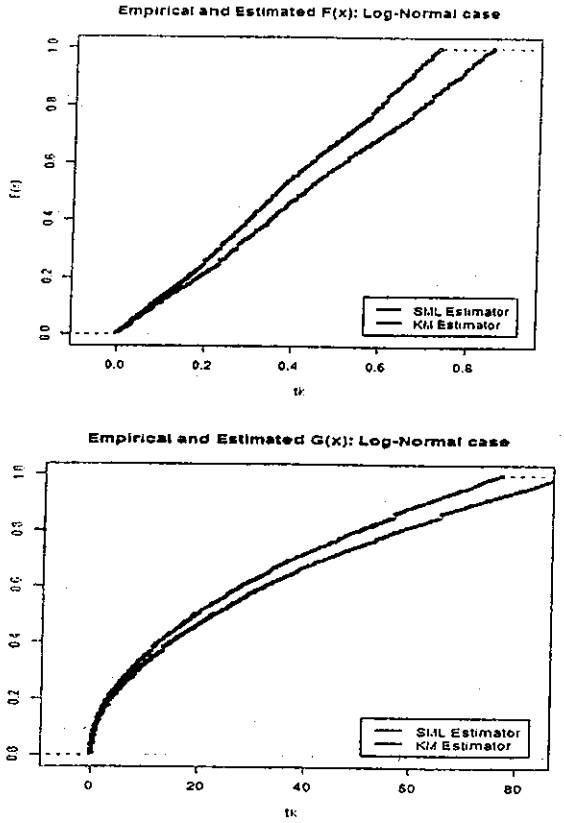
من الجدول السابق نستنتج أن هناك فرق بين المجموعتين محل التطبيق في حالة استخدام كلاً من التوزيع الأسي أو التوزيع اللوغاريتمي الطبيعي حيث أن قيمة الأحتمالية المحسوبة p-value في كلا الحالتين أقل من 5% وبالتالي لا نستطيع أن نقبل الفرض العدمي القائل بأن المجموعتين متساويتين في التأثير وعليه يمكننا القول بأن المجموعتين مختلفتين في التأثير باستخدام المعالجة المقترحة في حالة مراقبة من جهة اليمين (لكلا المجموعتين من الذكور والاناث).

الأشكال البيانية التالية توضح مقارنة بين تقدير $\hat{F}(x)$ وتقدير $\hat{G}(x)$ باستخدام مقدرات الأمكان الأعظم شبه المعلمي (SML) ومقدر Kolmogorov–Smirnov (KM). في شكل (١) يوضح تقدير كلاً من تقدير $\hat{F}(x)$ وتقدير $\hat{G}(x)$ في حالة توزيع الأسي بينما شكل (٢) يوضح تقدير كلاً من تقدير $\hat{F}(x)$ وتقدير $\hat{G}(x)$ في حالة توزيع اللوغاريتمي الطبيعي، وكما هو واضح من الأشكال البيانية تقارب طريقتي التقدير: الأمكان الأعظم شبه المعلمي (SML) ومقدر Kolmogorov–Smirnov (KM) وهو مطابق للنتائج النظرية التي تم الحصول عليها في جدول رقم (٢).

شكل (١): مقارنة بين تقدير $\hat{F}(x)$ وتقدير $\hat{G}(x)$ باستخدام مقدرات الأمكان الأعظم شبه المعلمي (SML) ومقدر Kolmogorov–Smirnov (KM) وذلك في حالة التوزيع الأسي



شكل (٢): مقارنة بين تقدير $\hat{F}(x)$ وتقدير $\hat{G}(x)$ باستخدام مقدرات الأمكان الأعظم شبه المعلمي (SML) ومقدر Kolmogorov–Smirnov (KM) وذلك في حالة التوزيع اللوغاريتمي الطبيعي



٧. النتائج:

يمكن إيجاز النتائج التى تم التوصل إليها فى هذا البحث فى النقاط التالية:

١- عند استخدام البيانات الحقيقية باستخدام البيانات المراقبة من اليمين فيوجد هناك فرق بين المجموعتين محل التطبيق في حالة استخدام التوزيع الأسي فإن قيمة p-value (0.01254) أي انها أقل من 5%

وبالتالي فإن المجموعتين مختلفتين في التأثير باستخدام المعالجة المقترحة في حالة المراقبة من جهة اليمين.

٢- عند التطبيق علي حالة توزيع الطبيعي اللوغاريتمي فإن قيمة p-value أيضاً (0.1281) أقل من 5% وبالتالي فإن المجموعتين (الذكور والاناث) مختلفتين في التأثير باستخدام المعالجة المقترحة في حالة المراقبة من جهة اليمين.

٣- تقارب دالتي الكثافة عند استخدام دالة الامكان الأعظم شبه المعلمي SML ومقدر كولوجروف سميرنوف KM في حالتي التوزيعين توزيع الأسي وتوزيع اللوغاريتمي الطبيعي مما يعني قوة الاختبار وأن النموذج المفترض ملائم لطبيعة البيانات في حالة وجود المراقبة من اليمين.

٨ . والتوصيات

بناءً على ماتوصلت إليه الدراسة من نتائج، توصي هذه الدراسة بالآتي:

١- استخدام طريقة تقدير أخري لتقدير معلمات نموذج DRM مع البيانات المراقبة المختلفة أو بيانات عادية.

٢- تطبيق نموذج نسبة الكثافة DRM في حالات التوزيعات الأخري لدالة الفشل مثل: توزيع جاما $h(x) = (x, log|x|)$ ، أو التوزيع الطبيعي $h(x) = (x, x^2)$.

٣- استخدام طريقة بيز لتقدير معلمات النموذج DRM ومقارنتها بنتائج تقديرات الامكان الأعظم.

المراجع

1) Anderson, J. A. (1979) "Multivariate logistic compounds". Biometrika, 66(1):17-26.

2) Anderson, T. W. and Darling, D. A. (1954), 'A Test of Goodness-of-Fit'. Journal of the American Statistical Association, 49(268), 765-769.

3) Cai, S. (2014). "On dual empirical likelihood inference under semiparametric density ratio models in the presence of multiple samples with applications to long term monitoring of lumber quality" (Doctoral dissertation, University of British Columbia).

4) Chakravarti, I. M., Laha, R. G., & Roy, J. (1967). "Handbook of methods of applied statistics." Wiley Series in Probability and Mathematical Statistics (USA) eng.

5) Chen, J., & Liu, Y. (2013). "Quantile and quantile-function estimations under density ratio model." The Annals of Statistics, 41(3), 1669-1692.

6) Cheng, K. F., & Chu, C. K. (2004). "Semi parametric density estimation under a two-sample density ratio model." Bernoulli, 10(4), 583-604.

7) Darling, D. A. (1957), 'The Kolmogorov-Smirnov, Cramer-von Mises Tests'. Ann. Math. Statist, 28(4), 823-838

8) Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 1-22.

9) Diao, G., & Ning, J. (2012). "Maximum likelihood estimation for semi parametric density ratio model." The international journal of biostatistics, 8(1).

10)Epstein, B., & Sobel, M. (1953)." Life testing." Journal of the American Statistical Association, 48(263), 486-502.

11)El Kassas, M., Alboraie, M., Omar, H., El Latif, Y. A., Algaber, M. A., El Tahan, A. & Doss, W. (2019). High success rates for the use of ombitasvir/paritaprevir/ritonavir containing regimens in treatment of naïve and experienced chronic hepatitis C genotype 4: Real world results. Journal of medical virology. DOI: 10.1002/jmv.25478, 2019;1-8

12)Fokianos, K. (2007). "Density ratio model selection. " Journal of Statistical Computation and Simulation, 77(9), 805-819

13)Zhang, A. G., & Chen, J. (2021). "Density ratio model with data-adaptive basis function." arXiv:2103.03445.

14)Zeng, D., Gao, F., & Lin, D. Y. (2017). "Maximum likelihood estimation for semi parametric regression models with multivariate interval-censored data." Biometrika, 104(3), 505-525.

15)Zhang, A. G., Zhu, G., & Chen, J. (2020). "Empirical Likelihood Ratio Test on quantiles under a Density Ratio Model." arXiv:2007.10586.

16)Zhuang, W. W., Hu, B. Y., & Chen, J. (2019). "Semi parametric inference for the dominance index under the density ratio model." Biometrika, 106(1), 229-241.