IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2026) 16 (2): 353-362

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521

University of Kirkuk Journal For Administrative and Economic Science

Al-Maadhedee Maan W. & Al-Sabaawe Younis M. Kh. Ministry of Higher Education and Scientific Research University of Sumer College of Administration and Economics Dhi Qar, Iraq. ***University of Kirkuk Journal For Administrative and Economic Science*** **(2026) 16 (2):353-362.**

Ministry of Higher Education and Scientific Research University of Sumer College of Administration and Economics Dhi Qar, Iraq

Iman Jamal Abdulalkadhim [1], Mohammed Muayad Shukur [2], Abbas Salem Kadhim [3]

[1,3] Ministry of Higher Education and Scientific Research-University of Sumer/College of Administration and Economics, Dhi Qar, Iraq

[2] Ministry of Education Dhi Qar Education Directorate Al-Rifa'i Education Department Dhi Qar, Iraq

Iman.Jamal@uos.edu.iq [1], MohammedMoayed.S@gmail.com [2], abbas.salem@uos.edu.iq [3]

Abstract: This research addresses the estimation of non-parametric extropy for length-biased data using an Expectation-Maximization (EM) iterative algorithm. Extropy serves as a complementary measure to entropy in measuring uncertainty, and its estimation faces significant challenges when data is biased. Therefore, the study aimed to develop an iterative framework (EM-KDE) that combines the Expectation-Maximization algorithm with kernel density estimation to enhance the accuracy of conclusions. The research relied on simulation results to analyze the impact of different methods used in estimating extropy. The findings revealed that the true extropy estimate was -0.1250, while the uncorrected estimate yielded a value of -0.0897, indicating a bias of 0.0354. In contrast, the corrected estimates, including the adjusted kernel density estimation (W-KDE) and the EM-KDE algorithm, both provided a value of -0.1046, reflecting a notable improvement and a reduction in deviation from the true value by 42.3%. Thus, the study recommends adopting corrected methods when analyzing biased data and conducting further research to explore new estimation techniques, in addition to developing analytical tools that facilitate the application of these methods in future research.

Keywords: Extropy, Length-bias, non-parametric estimation, Kernel Density Estimation (KDE), Expectation-Maximization (EM)

تحسين تقديرات الإكستروبي اللامعلمي للبيانات المتحيزة بالطول باستخدام خوارزمية تكرارية (EM)

م.م. ايمان جمال عبد الكاظم [١] ، م.م. محمد مؤيد شكر [٢]، م.م. عباس سالم كاظم [٣]

[١،٣] وزارة التعليم العالي والبحث العلمي / جامعة سومر / كلية الإدارة والاقتصاد، ذي قار، العراق

[٢] وزارة التربية /مديرية تربية ذي قار/قسم تربية الرفاعي، ذي قار، العراق

IRAQI
Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2026) 16 (2): 353-362

Iman.Jamal@uos.edu.iq [1], MohammedMoayed.S@gmail.com [2], abbas.salem@uos.edu.iq [3]

المستخلص: تناول البحث تقدير الإكستروبي اللامعلمي للبيانات المتحيزة بالطول باستخدام خوارزمية تكرارية (EM) حيث يعتبر الإكستروبي مقياساً مكملاً للانتروبي في قياس عدم اليقين، ويواجه تقديره تحديات كبيرة في حال وجود تحيز في البيانات. لذا هدفت الدراسة إلى تطوير إطار عمل تكراري (EM-KDE) يجمع بين خوارزمية توقع-تعظيم وتقدير كثافة النواة لتحسين دقة الاستنتاجات. حيث استند البحث إلى نتائج المحاكاة لتحليل تأثير الأساليب المختلفة المستخدمة في تقدير الإكستروبي. أظهرت النتائج أن تقدير الإكستروبي الحقيقي كان -٠,١٢٥٠، بينما أظهر التقدير غير المصحح قيمة -٠,٠٨٩٧، مما يشير إلى تحيز قدره ٠,٠٣٥٤. أما التقديرات المصححة، بما في ذلك تقدير الكثافة المصححة (W-KDE) وخوارزمية EM-KDE، فقدمت كلاهما قيمة -٠,١٠٤٦، مما يعكس تحسناً ملحوظاً وتقليلاً في الانحراف عن القيمة الحقيقية بنسبة ٤٢,٣%.لذا توصي الدراسة بتبني الأساليب المصححة عند تحليل البيانات المتحيزة، وإجراء دراسات إضافية لاستكشاف طرق تقدير جديدة، بالإضافة إلى تطوير أدوات تحليلية تسهل تطبيق هذه الأساليب في الأبحاث المستقبلية.

الكلمات المفتاحية: الإكستروبي، التحيز بالطول، التقدير غير المعلمي، تقدير كثافة النواة، خوارزمية توقع-تعظيم (EM)

Corresponding Author: E-mail: Iman.Jamal@uos.edu.iq

المقدمة

يُعتبر تقدير دالة التوزيع والكثافة الاحتمالية ركيزة أساسية في التحليل الإحصائي، وتقوم عليه معظم مقاييس عدم اليقين والمخاطر. لطالما كان الانتروبي(Entropy) ، الذي أسس له شوان، هو المقياس التقليدي. وفي الآونة الأخيرة، ظهر مقياس الإكستروبي (Extropy)كقياس مزدوج ومُكمِّل للانتروبي، حيث قدمه لاد وزملاؤه، واكتسب أهمية متزايدة في مجالات الموثوقية وتحليل البيانات. في المقابل، تواجه العديد من الدراسات تحديات في جمع البيانات، أبرزها ظاهرة التحيز في الاختيار (Selection Bias)، التي غالباً ما تأخذ شكل المعاينة المتحيزة بالطول (Length-Biased Sampling - LBS) تحدث هذه الظاهرة في الدراسات ذات الفترات الزمنية مثل فترات البطالة أو مدة بقاء المنتجات، أو مدة بقاء الانسان على قيد الحياة...الخ حيث تكون المشاهدات ذات المدد الأطول أكثر احتمالاً للرصد، وإن استخدام البيانات المتحيزة مباشرة لتقدير خصائص التوزيع الأصلي يؤدي إلى استنتاجات خاطئة ومتحيزة . لذا يُعد التقدير اللامعلمي للإكستروبي ($J(X)$) حلاً قوياً لتجنب الافتراضات التوزيعية، ولكنه يتطلب منهجية مبتكرة لتصحيح تحيز الطول. وهنا يبرز دور خوارزمية التوقع-التعظيم - (Expectation-Maximization (EM)، التي تتعامل مع البيانات الناقصة أو غير المكتملة وذلك لأنها قد تخلق إطاراً تكرارياً مثالياً لاستعادة التوزيع الأصلي غير المتحيز من البيانات المتحيزة بالطول. لذا وبسبب التحيز الحاصل بالمتوسط والبيانات برزت الحاجة إلى خوارزميةEM في الوقت الذي لا يمكن استخدام المقدِّرات الإحصائية القياسية (مثل مقدّر كثافة النواة (KDE)) بشكل مباشر. من هنا تأتي أهمية الخوارزميات التكرارية EMلتقدير التوزيع الحقيقي غير المتحيز f(t) استناداً إلى البيانات المتحيزة المرصودة. وذلك لتحسينها عن طريق التعامل مع البيانات الحقيقية غير المرصودة (أوقات الحياة القصيرة) كـ بيانات مفقودة(Missing Data) ، ومن ثم استخدام خوارزمية تكرارية (EM) للوصول إلى تقدير غير متحيز.

المبحث الأول: منهجية البحث

اولاً: مشكلة البحث Problem Statement

تُعد تقديرات الإكستروبي اللامعلميه أدوات أساسية في مجالات متنوعة، بما في ذلك التحليل الإحصائي الحيوي وهندسة الموثوقية، حيث يعتمد اتخاذ القرارات الحاسمة على الفهم الدقيق لتقديرات أزمنة البقاء على قيد الحياة. ومع ذلك، تواجه هذه التقديرات تحديات كبيرة عند تطبيقها على البيانات المتحيزة بالطول والتحيز الحاصل بمتوسط التوزيع، مما يؤدي إلى تحيز في تقدير الإكستروبي الحقيقي للتوزيع الأصلي. يرجع هذا التحيز إلى أن التقديرات اللامعلميه التقليدية، مثل تلك المعتمدة على مُقدِّر النواة المباشر، لا تأخذ في الاعتبار طبيعة التحيز الموجود في البيانات المتحيزة بالطول. لذا فأن هناك نقصاً ملحوظاً في وجود إطار تكراري متكامل يعتمد على خوارزمية EM ويمثل هذا النقص فجوة بحثية هامة، حيث أن دمج التقدير اللامعلمي لدالة الكثافة (Kernel Density Estimation - KDE) مع مقياس الإكستروبي ضمن إطار تكراري يمكن أن يحسن بشكل كبير من دقة وموثوقية تقديرات الإكستروبي.

ثانياً: أهداف البحث Research Objectives

يهدف البحث إلى:

١. **تطوير انموذج EM-KDE :** تطوير إطار عمل تكراري يدمج خوارزمية EM مع تقدير الكثافة باستخدام النواة(KDE) للحصول على تقدير مُصحَّح لدالة الكثافة الاحتمالية ($f(x)$) للتوزيع الأصلي غير المتحيز.

٢. **اشتقاق مُقدِّر الإكستروبي المُحسَّن :**اشتقاق وصياغة مُقدِّر لا معلمي جديد للإكستروبي ($\hat{J}_{EM}(X)$) يعتمد على دالة الكثافة المقدرة في كل تكرار من EM

٣. **دراسة الخصائص الإحصائية :**تحليل ودراسة خصائص التقارب (Convergence)والاتساق (Consistency) للمُقدِّر المقترح، ومقارنة أدائه (من حيث الانحياز ومتوسط مربعات الخطأ) مع المُقدِّرات اللامعلميه المنافسة في ظل التحيز بالطول والتبتير(LBD) .

٤. **التطبيق العملي :**تطبيق المنهجية المقترحة عن طريق المحاكاة خاضعة للتحيز بالطول، مثل بيانات مدة الحياة أو أوقات الانتظار.

ثالثاً: أهمية البحث وجانب الحداثة Significance and Novelty

تكمن حداثة هذا البحث وأهميته فيما يلي:

- الدمج المنهجي الثلاثي :يمثل البحث دمجاً مبتكراً بين نظرية المعلومات (الإكستروبي)، مشكلة المعاينة التحيز بالطول (LBD)، ومنهجية التعظيم التكرارية .(EM) هذا الدمج ينتج أداة تحليلية أقوى.
- تحسين الأداء :من المتوقع أن يوفر الإطار التكراري لخوارزمية EM مُقدِّراً أكثر استقراراً وأقل حساسية لاختيار عرض النطاق (Bandwidth) في تقدير النواة مقارنةً بالتقديرات اللامعلميه الصريحة الأخرى.
- التعامل مع التبتير (Censoring): المنهجية المقترحة باستخدام EM قادرة على التوسع للتعامل مع السيناريو الأكثر واقعية، وهو البيانات المتحيزة بالطول (LBD)، مما يزيد من قابلية التطبيق العملي للنتائج.

رابعاً: الاستعراض المرجعي Literature Review

- في عام ١٩٤٨، قدم العالم (Shannon) دراسة تأسيسية بعنوان "النظرية الرياضية للاتصال"، والتي وضع من خلالها حجر الأساس لعلم نظرية المعلومات .قدم البحث منهجية رياضية لقياس عدم اليقين في الأنظمة العشوائية عبر مفهوم الانتروبي (Entropy)، وأثبتت النتائج أن هذا المقياس هو الأفضل والوحيد القادر على كميّة متوسط المعلومات المفقودة أو "الارتباك" في توزيع الاحتمالات.
- في عام ١٩٥٨، قدم الباحثان (Kaplan & Meier) بحثاً درسوا فيه مقدراً لامعلمياً لحساب دالة البقاء من المشاهدات غير المكتملة(Incomplete Observations) ، والمعروف حالياً بمقدر "كابلان-ماير". ركزت الدراسة على كيفية التعامل مع البيانات المراقبة (Censored Data) دون فرض توزيع احتمالي محدد، وأثبتت النتائج أن هذا المقدر يوفر تقديرات ذات كفاءة عالية وغير متحيزة لدالة البقاء دون الحاجة لفرض توزيع احتمالي معين، متفوقاً بذلك على طرق الجداول الاكتوارية التقليدية التي كانت تفشل في استيعاب التبتير بشكل دقيق.
- في عام ١٩٧٨، أسس الباحثان (Patil & Rao) الإطار النظري للتوزيعات الموزونة (Weighted Distributions) لمعالجة ظاهرة انزياح الطول. قارنت الدراسة بين التوزيعات الأصلية والتوزيعات الموزونة(Size-biased) ، وأوضحت النتائج أن النماذج الموزونة هي الأفضل والوحيدة القادرة على تمثيل البيانات في حالات أخذ العينات غير العشوائي (دراسات الحياة البرية والأسر)، مما وفر الأساس الرياضي لتصحيح الانحياز في التقديرات اللاحقة.
- في عام ١٩٨٢، قدم الباحث (Vardi) مقدراً لامعلمياً للأرجحية القصوى (NPMLE) للتوزيع الأصلي عند وجود انزياح في الطول. قارن البحث بين طرق التقدير المباشرة والتقدير عبر الأوزان، وأثبتت النتائج أن مقدر (NPMLE) هو الأفضل في استعادة خصائص المجتمع الأصلي من العينة المنزاحة، كما تميز بكونه مقدراً متسقاً (Consistent) وقوياً إحصائياً في معالجة البيانات غير المكتملة.
- في عام ٢٠١١، قدم الباحثان (Gupta & Chen) دراسة معمقة لنظرية وتطبيقات خوارزمية EM في مجالات متعددة. استعرض الباحثون فاعلية الخوارزمية مقارنة بطرق الأخرى مثل(Newton-Raphson)، وأظهرت النتائج أن خوارزمية EMهي الأفضل من حيث الاستقرار العددي (Numerical Stability) وسهولة البرمجة، خاصة في النماذج التي تحتوي على متغيرات كامنة أو بيانات غير مكتملة، مما وسع نطاق استخدامها في معالجة الإشارات والإحصاء المتقدم.
- في عام ٢٠١٢، قدم الباحثان (Sunoj & Linu) بحثاً أدخلا فيه مقياساً معمماً يُعرف بـ "إنتروبي ريني التراكمي المتبقي الديناميكي". تناول البحث دراسة الخصائص الرياضية لهذا المقياس وتطبيقاته في النماذج الموزونة ونماذج التوازن (Equilibrium Models). توصلت الدراسة إلى أن هذا المقياس يوفر مرونة أكبر في تحليل البيانات التي تتغير خصائصها الإحصائية مع الزمن، خاصة عند التعامل مع توزيعات ذات ذيول ثقيلة.
- في عام ٢٠١٥، قدم الباحثون (Lad, Sanfilippo, & Agrò) مفهوم "الإكستروبي" (Extropy) "كبديل ومكمل ثنائي للانتروبي (Entropy) في قياس عدم اليقين. ناقش البحث الخصائص الرياضية لهذا المقياس في سياق التوزيعات الاحتمالية المستمرة، وأظهرت النتائج أن الإكستروبي يتفوق في قدرته على قياس التباعد بين دالة الكثافة الاحتمالية ودالة التوزيع التراكمي، كما أثبتوا فاعلية "الإكستروبي النسبي" كأداة قوية في تسجيل التوقعات المتسلسلة، مما فتح آفاقاً جديدة لاستخدامه في نماذج الموثوقية وتحليل البقاء.
- في عام ٢٠١٩، تناول الباحثان (Noughabi & Jarrahiferiz) تقدير دالة الإكستروبي باستخدام مقدرات قائمة على (Kernel-based estimators). قارنت الدراسة أداء المقدر المقترح مع مقدر الإكستروبي التجريبي (Empirical Extropy).وأظهرت نتائج المحاكاة أن المقدر المعتمد على (Kernel Estimator) "كان الأفضل والأكثر دقة، خاصة في

العينات الصغيرة والمتوسطة، حيث ساهم التنعيم في تقليل التحيز (Bias) الناتج عن فجوات البيانات مقارنة بالمقدر التجريبي الخام.

- في عام ٢٠٢٠، قدم الباحث (Chaubey) دراسة حول مقدر الانتروبي المعتمد على تقدير الكثافة اللامعلمي للبيانات غير السالبة .قارن البحث بين "مقدر (Adaptive Kernel)" والمقدرات الثابتة، وتبينت النتائج أن المقدر اللامعلمي الذي يستخدم "تحويلات (Kernel Transformations)" كان الأفضل في معالجة انحراف التوزيعات الطبية، حيث وفر تمثيلاً أدق لعدم اليقين مقارنة بالطرق المعلمية التي تفشل عند عدم تحقق فرضية التوزيع الطبيعي.
- في عام ٢٠٢١، قدم الباحثون (Shi et al) دراسة حول المقدر المعدل للبيانات التي تعاني من "انزياح الطول" و"التبتير اليميني" معاً. استخدم الباحثون التمثيل القوي (Strong Representation) للمقدر اللامعلمي، وأظهرت النتائج أن المنهجية المقترحة هي الأفضل لمواجهة التحديات المزدوجة (الانزياح والتبتير)، حيث وفرت دقة تنبؤية أعلى في تحليل أوقات البقاء مقارنة بالمقدرات التي تعالج كل مشكلة على حدة.
- في العام نفسه قدم الباحثون (Vaselabadi et al.) مقياس "الفار-إكستروبي" (Varextropy)باستخدام نماذج معدل المخاطرة النسبية (Proportional Hazard Rate Models) . قارنت الدراسة أداء هذا المقياس في ترتيب المتغيرات العشوائية، وأثبتت النتائج أن مقياس "الفار-إكستروبي المتبقي" هو الأفضل في تمييز الفروقات الدقيقة في ذيول التوزيعات مقارنة بمقاييس التشتت التقليدية، مما يجعله أداة متفوقة في (Stochastic Comparisons).
- في عام ٢٠٢٤، قدم الباحثون (Qubbaj et al.) بحثاً درسوا فية تقدير الإكستروبي والإنتروبي تحت نظام المراقبة المرحلية من النوع الأول (Type-I interval censoring). قارنت الدراسة بين تقدير الإمكان الأعظم (MLE) وتقدير "بيز " (Bayesian Estimation) تحت دوال خسارة مختلفة. وأظهرت النتائج أن تقدير بيز باستخدام "المعاين المسبق غير المعلوماتي (Non-informative prior)" كان الأفضل في تقليل الأخطاء المعيارية وتوفير فترات ثقة أكثر دقة عند مقارنتها بطريقة MLE التقليدية في ظروف البيانات المراقبة مرحلياً.
- في دراسة حديثة (٢٠٢٥) قدم للباحثين (Nair & Sathar) ، بحثاً اقترحوا فيه مقدرات لامعلمية لمقاييس الإكستروبي المعتمدة على بيانات منزاحة الطول. قارن الباحثون بين مقدر "تجارب مونت كارلو (Monte Carlo simulation) " والمقدرات التحليلية المقترحة، وأكدت النتائج أن المقدرات اللامعلميه المقترحة كانت الأفضل لامتلاكها خصائص الاتساق (Consistency)والاعتدال المقاربي(Asymptotic Normality) ، حيث أظهرت كفاءة عالية في التعامل مع البيانات الحقيقية للملوثات البحرية مقارنة بالمقدرات التقليدية التي لا تأخذ انزياح الطول في الحسبان.

المبحث الثاني: الجانب النظري

تمهيد:

يعد تحليل البيانات المتحيزة بالطول تحدياً جوهرياً في ميدان الإحصاء، خاصة في ظل الاعتماد المتزايد على البيانات في مختلف المجالات العلمية. حيث يهدف هذا الجانب من البحث إلى استكشاف مفهوم ما بين البيانات المبتورة (غير المكتملة) وبيانات الدراسة المتحيزة بالطول، سواء كانت بيانات مبتورة أو متحيزة بالطول، وكيف يؤثر ذلك على دقة التحليل الإحصائي. علاوة على ذلك، تناول مفهوم الإكستروبي كأداة لقياس عدم اليقين في التوزيعات الاحتمالية، وكيفية استخدام تقدير كثافة النواة (KDE) أي التقدير الغير مصحح وكذلك طرائق التقدير المصححة لذا يهدف إلى تزويد الباحثين بفهم عميق للتحديات التي تواجه الإحصائيين والعلماء في تحليل البيانات المتحيزة بالطول، وتسليط الضوء على الأساليب الحديثة المستخدمة لتقليل هذه التحيزات، مما يسهم في تحسين دقة النتائج والتفسيرات المستخلصة من البيانات.

اولاً: البيانات غير المكتملة (Incomplete Data)

تُعد البيانات غير المكتملة تحدياً رئيساً في التحليل الإحصائي، وتظهر في مواقف مختلفة وهي تكون على النحو الاتي:

١-البيانات المبتورة (Censored Data):

هي نوع من البيانات غير المكتملة تُرصد فيها الأحداث فقط إذا كانت أوقات ظهورها أقدم من وقت الرصد. في هذه الحالة، تُفقد المعلومات المتعلقة بالأحداث التي تحدث قبل نقطة زمنية معينة، مما يؤدي إلى نقص في البيانات المتاحة للتحليل. حيث تحدث هذه الظاهرة بشكل خاص في دراسات البقاء، حيث قد تنتهي الدراسة قبل تسجيل حدوث الحدث (مثل وفاة المريض)، مما يجعل المشاهدة جزئية. أي انه يمكن أن تؤدي النقاط الزمنية المفقودة إلى تحيز في التقديرات، حيث تُفقد المعلومات حول الأحداث التي تحدث مبكراً.

٢-البيانات المتحيزة بالطول (Length-Biased Data)

يعود اكتشاف مفهوم البيانات المتحيزة بالطول إلى منتصف القرن العشرين، حيث ساهم عدد من العلماء في صياغته وتطويره. يُعتبر C. R. Rao من أوائل الإحصائيين الذين قدموا الأساس النظري لفهم التوزيعات الموزونة (Weighted Distributions)، [18]وهي الإطار الذي تُعتبر البيانات المتحيزة بالطول حالة خاصة منه. أن احتمال اختيار وحدة من المشاهدات يتناسب مع قيمة المتغير العشوائي نفسه، مما أدى إلى زيادة فهمنا لكيفية تأثير طول المدة على عملية أخذ العينات. وفي الستينات، برزت أهمية البيانات المتحيزة بالطول بشكل خاص في مجالات الوبائيات (Epidemiology) وتجديد العمليات

ساهم كل من (Renewal Processes). D. R. Cox و H. D. Miller في توضيح مشكلة التحيز الزمني [4]، حيث أظهرت دراساتهم كيف أن الأحداث الأطول كانت أكثر احتمالًا للاكتشاف أثناء التحليلات الوبائية. بدورهم قاموا بتعزيز هذا المفهوم من خلال تطوير المقدِّر الأمثل للاحتمالية القصوى غير المعلمية - (Nonparametric Maximum Likelihood Estimator (NPMLE)، الذي يُستخدم لتوزيع وقت الحياة الحقيقي بناءً على العينات المتحيزة بالطول والمبتورة نحو اليسار. يُعتبر عمل فاردي حجر الزاوية في المعالجة الإحصائية غير المعلمية للبيانات المتحيزة. لذا تُعد البيانات المتحيزة بالطول مشكلة أساسية في الإحصاء، حيث تنشأ عندما تكون عملية أخذ العينات غير عشوائية، مما يؤدي إلى أن الأحداث أو الفترات الزمنية الأطول تكون أكثر تمثيلًا في العينة المرصودة مقارنة بالأحداث القصيرة. بمعنى تكون البيانات المرصودة أطول في المتوسط من البيانات الحقيقية في المجتمع الأصلي (مما يؤدي إلى تضخيم في تقديرات وقت الحياة أو البقاء) ويمكن التعبير عنه بالصيغة الرياضية الاتية:

افترض ان X هو متغير عشوائي لوقت الحياة (المدة الحقيقية غير متحيزة) وله دالة كثافة $f(x)$ بمتوسط $\mu = E[X]$.
اذا كانت عملية اخذ العينات تختار X باحتمالية تتناسب مع X نفسها، فان التوزيع الناتج (توزيع متحيز بالطول) $f^L(x)$ يكون:

$$f^L(x) = \frac{x\,.f(x)}{\mu} \qquad \ldots\ldots\ldots\ldots\ldots (1)$$

حيث ان μ هو متوسط المتغير الاصلي X.

- أي انه في البيانات المتحيزة بالطول، يؤدي التحيز في أخذ العينات إلى تضخيم تقديرات متوسط الزمن.[12,25]

٣- الإكستروبي Extropy $J(X)$

الإكستروبي هو مقياس الازدواجية (Complementary Dual) [11,17] اما الانتروبي (Entropy) ويستخدم لقياس عدم اليقين في توزيع الاحتمالات[19,2,1] . بالنسبة لمتغير عشوائي مستمر X يمتلك دالة كثافة الاحتمال $f(x)$ ، يُعرف الإكستروبي $J(X)$ بالمعادلة التالية:

$$J(X) = -0.5 \int_{R_X} f^2(x)dx \qquad \ldots\ldots\ldots\ldots (2)$$

حيث R_X هو مجال دعم المتغير العشوائي X. يُقدر الإكستروبي باستخدام تقدير دالة الكثافة $\hat{f}(x)$ على شبكة نقاط X_{grid} وباستخدام طريقة التكامل العددي مثل قاعدة شبه المنحرف (trapz):

$$\hat{J}(x) = -0.5 \int (\hat{f}(x))^2\, dx \approx -0.5 \cdot trapz\,(X_{grid}, (\hat{f}_{hat})^2) \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

٤- مقدّر كثافة النواة (Kernel Density Estimator - KDE)

وهي تقنية غير معلميه تُستخدم لتقدير دالة كثافة الاحتمال لمتغير عشوائي بشكل سلس ومستمر، بدلاً من استخدام المدرج التكراري (Histogram). يُستخدم مُقدِّر النواة (Kernel Estimator) لتقدير دالة كثافة الاحتمال $f(x)$ بطريقة غير معلميه ويعتمد هذا المُقدِّر على تحديد "نافذة" (window) حول كل نقطة بيانات، واستخدام دالة نواة (Kernel function) لوضع وزن أكبر للملاحظات القريبة من النقطة التي يتم التقدير عندها. مُقدِّر كثافة النواة $\widehat{f_n}(x)$ لدالة كثافة الاحتمال $f(x)$ يُعطى بالصيغة:

$$\widehat{f_n}(x) = \frac{1}{N}\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-X_i}{h}\right) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

طريقة التقدير الغير مصحح (Naive Estimation)

تطبق هذه الطريقة تقدير KDE المعتاد مباشرةً على البيانات المتحيزة X دون تصحيح، مما يؤدي إلى تقديرات متحيزة للإكستروبي.

- **الأوزان:** أوزان متساوية

$$w_i = \frac{1}{N}$$

- **تقدير دالة الكثافة** $\hat{f}_{\text{Naïve}}(x)$

$$\hat{f}_{\text{Naive}}(x) = \sum_{i=1}^{N} w_i K_h(x - X_i) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

N تمثل حجم العينة.
h يمثل معامل التنعيم
K دالة النواة *kernel function*
$\hat{f}_{Naive}(x)$ هي تقدير لدالة الكثافة

• **تقدير الإكستروبي الغير مصحح :**

$$\hat{J}_{\text{Naive}} = -0.5\int \left(\hat{f}_{\text{Naive}}(x)\right)^2 dx \quad \text{..........} (6)$$

٥- طريقة التقدير المصحح غير التكراري W-KDE

تعالج هذه الطريقة التحيز باستخدام أوزان غير متساوية ومُعايرة للعينات المرصودة، مستمدة من مفهوم تقدير الاحتمالية القصوى اللامعلميه (NPMLE) .

• **الأوزان المُصحَحة** w_i

$$w_i = \frac{1/x_i}{\sum_{j=1}^{N}(1/x_j)} \quad \text{....................} (7)$$

• **تقدير دالة الكثافة** $\hat{f}_{\text{Corrected}}(x)$

$$\hat{f}_{\text{Corrected}}(x) = \sum_{i=1}^{N} w_i \cdot \frac{1}{h} K\left(\frac{x-X_i}{h}\right) \quad \text{................} (8)$$

• **تقدير الإكستروبي المصحح:**

$$\hat{J}_{\text{Corrected}} = -0.5\int \left(\hat{f}_{\text{Corrected}}(x)\right)^2 dx \quad \text{................} (9)$$

٦- طريقة التقدير التكراري EM-KDE

تستخدم هذه الطريقة خوارزمية توقع-تعظيم (Expectation-Maximization - EM) لتقدير دالة الكثافة الأصلية $f(x)$.[8]
مراحل خوارزمية (EM-KDE) المستخدمة في البحث:

أ- المعلمة البدائية: ($\hat{\mu}_X^{(0)}$) تعيين قيمة أولية لمتوسط المتغير الأصلي $E[X]$.

ب- مرحلة التوقع (E-Step): حساب الأوزان w_i:

$$w_i = \frac{\frac{1}{X_i}}{\sum_{j=1}^{N}\left(\frac{1}{X_j}\right)} \quad \text{... ... ...} (10)$$

ت- مرحلة التعظيم:(M-Step)

- **تقدير دالة الكثافة M-Step 1**

حساب تقدير دالة الكثافة $\hat{f}_{\text{EM}}^{(\text{current})}(x)$ باستخدام الأوزان w_i

$$\hat{f}_{\text{EM}}^{(\text{current})}(x) = \sum_{i=1}^{N} w_i \cdot \frac{1}{h} K\left(\frac{x-X_i}{h}\right) \text{... ... ...} (11)$$

- **تقدير الإكستروبي المصحح باستخدام خوارزمية:EM M-Step 2**

$$\hat{J}_{\text{EM}}^{(\text{current})} = -0.5\int \left(\hat{f}_{\text{EM}}^{(\text{current})}(x)\right)^2 dx \text{ ....} (12)$$

- **تحديث معيار التقارب: M-Step 3**

تحديث قيمة متوسط المتغير الأصلي $\hat{\mu}_X$ باستخدام الأوزان الحالية w_i

$$\hat{\mu}_X^{(\text{new})} = \sum_{i=1}^{N} X_i \cdot w_i \quad \text{...............} (13)$$

ث- التقارب: (Convergence Check) تتوقف العملية التكرارية عندما يكون التغير المطلق في قيمة $\hat{\mu}_X$ أقل من حد التسامح

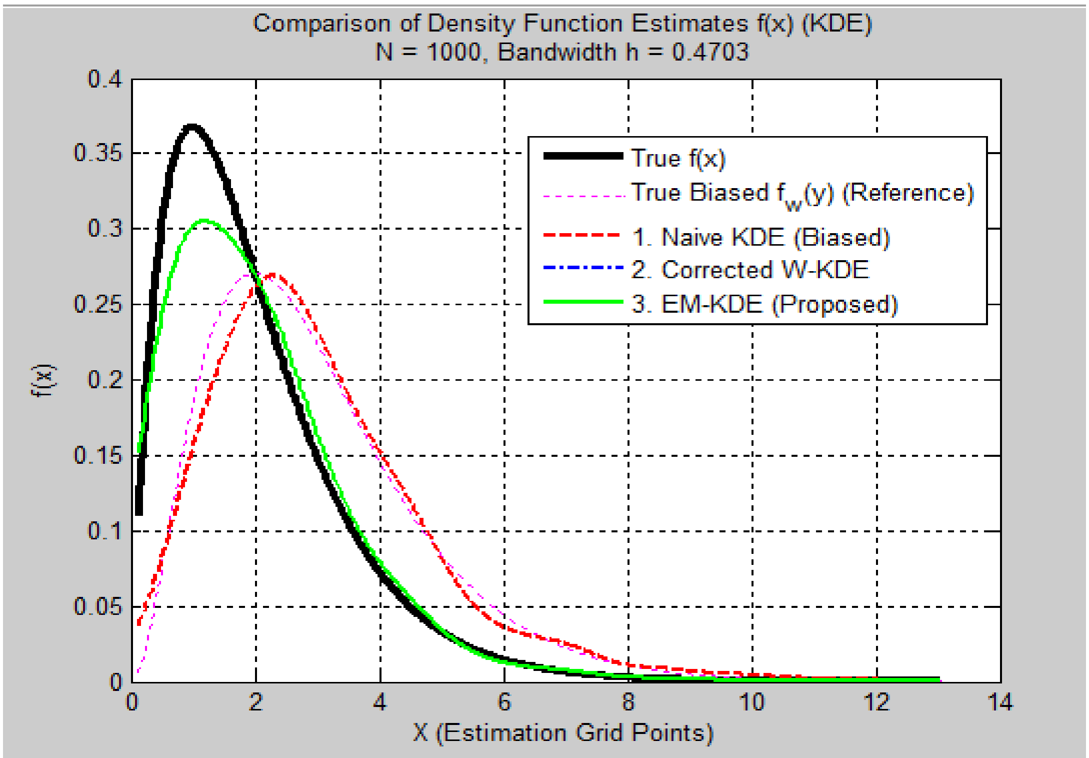
$$|\, \hat{\mu}_X^{(\text{new})} - \hat{\mu}_X^{(\text{old})} \,| < \text{Tolerance} \quad \text{..........} (14)$$

قيمة $\hat{J}_{\text{EM}}$هي القيمة النهائية $\hat{J}_{\text{EM}}^{(\text{current})}$ عند تحقيق التقارب.

المبحث الثالث: الجانب العملي

تمهيد

مع التقدم المستمر في مجالات الإحصاء وتحليل البيانات، تبرز أهمية تقديرات دالة الكثافة كأداة أساسية لفهم التوزيعات المعقدة للبيانات المتحيزة. ففي هذا السياق، يتناول الجانب العملي من البحث تقييم مجموعة منهجية من الأساليب المستخدمة في تقدير دالة الكثافة، مع التركيز على كلٍ من الطرائق غير المصححة والطرائق المصححة المقترحة. تم الاعتماد على تقنيات المحاكاة لتحليل هذه الأساليب وتقييم أدائها في ظروف مختلفة باستخدام برنامج MATLAB-V-21، ومن خلال التقييم الشامل يؤدي الى توفير دليل مفيد للباحثين والممارسين في مجال تحليل البيانات المتحيزة بالطول، مما يمكنهم من اتخاذ قرارات مستنيرة مبنية على نتائج موثوقة تعكس التوزيعات الواقعية للبيانات. وكما موضح في ادناه:



الشكل (١): مقارنة تقديرات دالة الكثافة للطرائق المختلفة

الرسم البياني اعلاه يظهر مقارنة بين تقديرات دالة الكثافة الناتجة عن عدة طرق مختلفة. يشتمل على أربع منحنيات رئيسة، كل منها يمثل تقديراً مختلفاً لدالة الكثافة، وهذا بدوره يعكس تأثير المنهجية المتبعة على دقة النتائج وكلاتي:

١. الدالة الحقيقية للكثافة:(True f(x))

يمثل هذا المنحنى التوزيع الفعلي للبيانات ويعد هذا الأساس الذي تُقارن به التقديرات الأخرى ويكون بشكل مثالي للشكل الذي نود التوصل إليه في تقديراتنا.

٢. الدالة المتحيزة الحقيقية (True Biased f(x)) :

هذا المنحنى يمثل تقديراً لكثافة البيانات المتحيزة حيث يظهر بشكل منقطع، ويعكس عدم الدقة في التقديرات. أي ان هذا التحيز في المنحنى لا يعبر عن الشكل الحقيقي للتوزيع بشكل مناسب.

٣. تقدير الكثافة الغير مصحح (Naive KDE) :

يظهر هذا المنحنى تحيزاً ملموساً، مما يدل على أنه يبتعد عن الشكل الحقيقي لدالة الكثافة حيث يعكس فشل الطرق الغير مصححه في التعرف على الشكل المعقد للتوزيع، مما يؤدي إلى نتائج غير دقيقة.

٤. تقديرات الكثافة المصححة (Corrected W-KDE) :

يمثل هذا المنحنى تقديراً محسّناً، حيث تم إدخال تصحيحات لمعالجة التحيز مما يظهر تحسناً ملحوظاً في دقة التقدير، كونه يقترب من الشكل الحقيقي لدالة التوزيع.

٥. التقدير باستخدام خوارزمية تكرارية EM-KDE :

يُظهر هذا المنحنى تحسينات كبيرة مقارنة بالطريقة الغير مصححه (مقدر كثافة النواة الغير مصححه)، مما يعكس الكفاءة العالية لخوارزمية التوقع-التعظيم (EM) في التعامل مع البيانات المتحيزة بالطول. حيث يضيف هذا التقدير دقة أكبر من التقديرات السابقة.

لذا يوضح الرسم ان التقنيات المصححة EM-KDE تظهر تحسينات كبيرة في النتائج. لذا يمكن أن نستنتج من الرسم أن استخدام تقديرات مصححة يُعتبر أفضل عند التعامل مع البيانات التي تكون متحيزة بالطول، حيث تعزز هذه الطرق دقة النتائج وتمنحنا فهماً أوضح للتوزيع الفعلي للبيانات. بالتالي، فإن اعتماد نماذج تحليل متقدمة يزيد من موثوقية الاستنتاجات، ويساعد في اتخاذ قرارات مدعومة بالبيانات.

جدول (١): مقارنة تقديرات الإكستروبي ومؤشرات الأداء

Method	Estimated Extropy	Bias	Squared Error
True J(X)	-0.1250	0	0
Naive J_hat	-0.0897	0.0354	0.0013
Corrected J_hat	-0.1046	0.0204	0.0004
EM-KDE J_hat	-0.1046	0.0204	0.0004

يوضح جدول (١) مقارنة تقديرات الإكستروبي ومؤشرات الأداء، بما في ذلك التقدير الحقيقي(True J(X)) ، والتقدير الغير مصحح (Naive J_hat)، بالإضافة إلى التقديرات المصححة التي تم الحصول عليها من خوارزمية التوقع-التعظيم (EM-KDE J_hat) . تمثل القيمة (-0.1250) للإكستروبي الحقيقي الأساس الذي تُقارن به التقديرات الأخرى. حيث يظهر التقدير الغير مصحح قيمة تساوي (-0.0897)، وهو أقل من القيمة الحقيقية وهذا يشير إلى أن استخدام هذا التقدير، دون معالجة التحيز الموجود في البيانات، يؤدي إلى استنتاجات غير دقيقة. فعلى الرغم من أن هذا التقدير يمكن أن يكون سريعاً وسهل الاستخدام، إلا أنه يعكس عدم القدرة على التعرف على التحيز في البيانات المتحيزة بالطول. أمّا بالنسبة للتقديرات المصححة، فقد قدمت القيم المتساوية (-0.1046) لكل من تقدير الإكستروبي المصحح (Corrected J_hat) وتقدير الإكستروبي باستخدام خوارزمية EM (EM-KDE J_hat). تشير هذه النتائج إلى تحسين كبير في دقة التقدير، حيث أخذت هذه الطرق في الاعتبار التحيز الموجود في البيانات، مما أدى إلى تقليل الانحراف عن القيمة الحقيقية بشكل ملحوظ. وعند النظر إلى مؤشر التحيز، نجد أن التقدير الغير مصحح يعكس تحيزاً ملموساً يبلغ ٠,٠٣٥٤، مما يعزز فكرة أنه يعاني من عدم دقة كافية. بالمقابل، يحافظ كل من التقديرين المصححين على مستوى تحيز أقل بكثير، حيث يبلغ ٠,٠٢٠٤. هذه النتيجة تعكس فعالية الأساليب المستخدمة لمعالجة التحيز، مما يشير إلى استنتاجات أكثر موثوقية.

أما بالنسبة لقياس الخطأ المربّع، فقد جاءت نتيجة التقدير الغير مصحح بأعلى مستوى بين جميع الطرق، حيث سجلت ٠,٠٠١٣. وتعبر هذه القيمة عن عدم الدقة وعدم موثوقية التقدير الغير مصحح. في المقابل، نجحت التقديرات المصححة في تقليل الخطأ المربّع إلى ٠,٠٠٠٤، مما يعكس تحسناً ملحوظاً ونتائج أكثر دقة. لذا تشير هذه النتائج إلى الفجوات الكبيرة في الدقة عند الاعتماد على التقديرات الغير مصححه للإكستروبي. علاوة على ذلك، يتضح من النتائج المصححة أن دمج خوارزمية التوقع-التعظيم (EM)مع تقدير كثافة النواة (KDE) يجعل من الممكن استعادة تقديرات دقيقة، متناولة التحيز الموجود في البيانات المتحيزة بالطول. هذه الدراسة تدعم أهمية استخدام أساليب محسّنة في التحليل الإحصائي، مما يمكن الباحثين من الحصول على نتائج أكثر موثوقية واتساقًا في سياق البيانات المعقدة.

جدول (٢): تقديرات قيم الإكستروبي

Naive Extropy Estimator J_Naive	**-0.0897**
Corrected Extropy Estimator J_Corrected	-0.1046
Iterative Extropy Estimator J_EM	-0.1046

يوضح جدول (٢) تقديرات الإكستروبي باستخدام أساليب مختلفة، مما يمنحنا فهماً أفضل لكيفية تأثير كل طريقة على النتائج. يظهر التقدير الغير مصحح (Naive Extropy Estimator J_Naive) قيمة قدرها (-0.0897). وتبين هذه القيمة أن التقدير الغير مصحح، بسبب عدم مراعاته للتأثيرات الناتجة عن التحيز، يؤدي إلى تقدير منخفض بشكل ملحوظ مقارنة بالقيمة الحقيقية. أما تقدير الإكستروبي المصحح (Corrected Extropy Estimator J_Corrected) فقد أظهر قيمة (-0.1046)، مما يدل على تحسين ملحوظ في الدقة بفضل الأدوات التصحيحية المستخدمة. وبالمثل، عكست قيمة تقدير الإكستروبي التكراري باستخدام خوارزمية EM (Iterative Extropy Estimator J_EM) نفس القيمة (-0.1046)، مما يعكس النجاح في استعادة تقديرات دقيقة من البيانات المتحيزة بالطول. لذا تتضح من هذه النتائج وجود فجوات كبيرة في الدقة التي تنجم عن الاعتماد على تقديرات غير مصححه للإكستروبي. بينما تقدم الأساليب المصححة، بما في ذلك استخدام خوارزمية التوقع-التعظيم (EM)، تقديرات محسّنة تأخذ بعين الاعتبار التحيز الموجود في البيانات. وبالتالي، فإن هذه الدراسة تؤكد على أهمية تطبيق أساليب محسّنة في التحليل الإحصائي لضمان الحصول على نتائج أكثر موثوقية ودقة، خصوصاً عند التعامل مع بيانات معقدة ومتنوعة.

الاستنتاجات والتوصيات

اولاً: الاستنتاجات

١. يوضح الرسم البياني الفروق الكبيرة في دقة تقديرات دالة الكثافة وفقاً للأساليب المستخدمة. حيث تُظهر الطريقة الغير المصححة تحيزاً ملحوظاً وانحرافاً عن الشكل الحقيقي لدالة الكثافة، مما يؤثر سلباً على صحة النتائج. بينما تتميز منحنيات الطرائق المصححة باقترابها من التوزيع الحقيقي للبيانات، مما يعكس فعالية النهج التصحيحي في معالجة العيوب. كما يظهر الرسم أن استخدام خوارزمية التوقع-التعظيم (EM-KDE) قدم تحسينات ملحوظة في الدقة، حيث ساعدت الطرائق المصححة في تقليل الانحراف عن القيمة الحقيقية، مما يعكس القدرة في التعرف على الهيكليات الخفية في البيانات.

٢. **تحقيق دقة تقدير الإكستروبي:**
أظهرت الدراسة أن القيمة الحقيقية للإكستروبي، والتي بلغت 0.1250-، تعتبر مرجعاً أساسياً للتقديرات الأخرى. وفي المقابل، سجل التقدير غير المصحح قيمة 0.0897-، مما يعني أنه انحرف بنحو (0.0353) عن القيمة الحقيقية، وهذا يشير إلى عدم دقة هذا التقدير.

٣. **مستوى التحيز:**
أظهر التقدير غير المصحح تحيزاً قدره 0.0354، بينما تمكنت الطرائق المصححة مثل تقدير الإكستروبي المصحح وخوارزمية التوقع-التعظيم (EM-KDE) من الحفاظ على مستوى تحيز أقل، قدّر بـ 0.0204 يعكس هذا التحسن بمقدار 42.3%، ما يشير إلى فعالية الأساليب المصححة في تقليل التحيز.

٤. **خطأ التقدير المربّع:**
أظهر التقدير غير المصحح أعلى مستوى من الخطأ المربّع، حيث بلغ 0.0013، مما يدل على قلة موثوقية هذا التقدير. بالمقابل، حققت الطرائق المصححة خطأ مربّعاً لم يتجاوز 0.0004، مستعرضةً تحسناً ملحوظاً بنسبة 69.2%، مما يعزز دقة النتائج المستخلصة.

٥. **تحسن تقدير الإكستروبي:**
أظهرت التقديرات المصححة قيمة قدرها 0.1046-، مما يعكس تحسناً واضحاً في دقة التقدير. وان هذه القيم تشير إلى أن الأساليب المصححة قد ساعدت في تقليل الانحراف عن القيمة الحقيقية.

٦. **كفاءة الخوارزمية**
كشفت النتائج أن خوارزمية التوقع-التعظيم (EM-KDE) قد أظهرت نفس قيمة تقدير الإكستروبي المحسّن(0.1046-) ، مما يبرز كفاءتها العالية في تحسين دقة التقدير والتعامل مع البيانات المتحيزة بالطول. بناءً على هذه النتائج، يتضح أن استخدام الأساليب المصححة في التقدير الإحصائي يعزز دقة النتائج وموثوقيتها، مما يعزز القدرة على التحليل المتقدم للبيانات المتحيزة بالطول.

ثانياً: التوصيات

١. **استخدام الأساليب المصححة :**ينصح الباحثون والممارسون بتبني الطرائق المصححة لتحليل البيانات، خاصة في حالات البيانات المتحيزة أو المعقدة، لضمان نتائج موثوقة.
٢. **إجراء دراسات إضافية :**ينبغي إجراء دراسات مستقبلية تستكشف أساليب تقدير أخرى غير مأخوذة في هذا البحث، بما في ذلك الاعتماد على تقنيات تعلم الآلة لتحسين دقة التقديرات.
٣. **تطوير أدوات تحليل متقدمة :**من المهم تطوير أدوات برمجية تسهل تطبيق الأساليب المصححة وتوفر واجهات مستخدم تجعل من السهل إدماجها في أبحاث وتحليلات البيانات.

References

1. Buono, F., Kamari, O., & Longobardi, M. (2023). Interval extropy and weighted interval extropy. *Ricerche di Matematica*, 72(1), 283–298.
2. Chaubey, Y. P. (2020). On an entropy estimator based on a non-parametric density estimator for non-negative data. *Journal of Statistical Distributions and Applications*, 7(2).
3. Chen, Y. C. (2017). A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1), 161–187.
4. Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. Methuen.
5. Etikan, I. (2017). The Kaplan Meier estimate in survival analysis. *Biometrics & Biostatistics International Journal*, 5(2), 55–59.
6. Gijbels, I. (2011). Kaplan-Meier estimator. In *International Encyclopedia of Statistical Science* (pp. 709–710). Springer, Berlin, Heidelberg.
7. Goel, M., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, 1(4), 274–278.
8. Gupta, M. R., & Chen, Y. (2011). Theory and use of the EM algorithm. *Foundations and Trends in Signal Processing*, 4(3), 223–296.
9. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481.
10. Krishnan, A. S., Sunoj, S. M., & Sankaran, P. G. (2020). Some reliability properties of extropy and its related measures using quantile function. *Statistica*, 80(4), 413–437.
11. Lad, F., Sanfilippo, G., & Agrò, G. (2015). Extropy: Complementary dual of entropy. *Statistical Science*, 30(1), 40–58.
12. Nair, R. D., & Sathar, E. I. A. (2025). Nonparametric estimation of extropy-related measures with length-biased data. *Statistics and Applications*, 23(1), 321–334.
13. Nanda, A. K., Sankaran, P. G., & Sunoj, S. M. (2014). Rényi's residual entropy: A quantile approach. *Statistics & Probability Letters*, 85, 114–121.
14. Njomen, D. A. N., & Ngatchou, J. W. (2018). Consistency of the Kaplan-Meier estimator of the survival function in competing risks. *The Open Statistics & Probability Journal*, 9(1), 1–17.
15. Noughabi, H. A., & Jarrahiferiz, J. (2019). On the estimation of extropy. *Journal of Nonparametric Statistics*, 31(1), 88–99.
16. Patil, G. P., & Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2), 179–189.
17. Qubbaj, H. H., Bayoud, H. A., & Hilow, H. M. (2024). Extropy and entropy estimation based on progressive Type-I interval censoring. *Statistics in Transition New Series*, 25(3), 83–102.
18. Rao, C. R. (1985). Weighted distributions: A review and some recent work. In *A Celebration of Statistics* (pp. 543–568). Springer, New York, NY.
19. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.
20. Shi, J., Chen, X., & Zhou, Y. (2021). The strong representation for the nonparametric estimator of length-biased and right-censored data. *arXiv preprint arXiv:1409.4642v2*.٩
21. Sunoj, S. M., & Linu, M. N. (2012). Dynamic cumulative residual Renyi's entropy. *Statistics*, 46(1), 41–56.
22. Sunoj, S. M., & Sankaran, P. G. (2012). Quantile based entropy function. *Statistics & Probability Letters*, 82(6), 1049–1053.
23. Tahmasebi, S., & Toomaj, A. (2020). On negative cumulative extropy with applications. *Communications in Statistics - Theory and Methods*.
24. Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10(2), 616–620.
25. Vaselabadi, N. M., Tahmasebi, S., Kazemi, M. R., & Buono, F. (2021). Results on varextropy measure of random variables. *Entropy*, 23(3), 356.