أخطاء التسجيل في سلاسل ماركوف
دراسة تطبيقية عن منتجات الألبان

بحث مقدم من قبل

مهدي محمد البياع

مدرس مساعد

مركز بحوث السوق وحماية المستهلك

جامعة بغداد

٢٠٠٣

أخطاء التسجيل في سلاسل ماركوف /دراسة تطبيقية

بحث مقدم من قبل

م.م.مهدي محمد البياع

مركز بحوث السوق وحماية المستهلك /جامعة بغداد

المستخلص

تعد فكرة الأخطاء بشكلٍ عام إحدى أهم الأفكار المثيرة للجدل سواء أكان الأمر بالنسبة للإحصائيين أو للباحثين من شتى المجالات الأخرى والذين يتعاملون مع هذه الفكرة ،حيث يعتقد البعض بأن موضوع الأخطاء هو من الموضوعات الشائكة والمعقدة ومن هنا برزت فكرة هذا البحث والتي تحاول إن تفسر فكرة الأخطاء غير العشوائية وبشكلٍ مبسط وبيان الفرق فيما بينها وبين الأخطاء العشوائية هذا وقد تم التطرق إلى واحدة من أهم وأخطر أنواع الأخطاء غير العشوائية ألا وهي أخطاء التسجيل وشرح إمكانية تقديرها والتنبؤ بمسارها ومقدارها وذلك في واحدةٍ من أهم أنواع التحاليل الإحصائية ألا وهي سلاسل ماركوف (*Markov Chain*) وبيان مدى تأثير وجود أخطاء التسجيل على قيم عناصر المصفوفة الانتقالية (*Transition Matrix*) وذلك عند كلاً من المدى البعيد (*Long Run*) وحالة الثبات والاستقرارية (*Stationary*) . أي أن مجال اهتمام هذا البحث انصب على التنبؤ بمقدار الأخطاء غير العشوائية في مصفوفة الاحتمالات الانتقالية عند الخطوة n وهل سيشكل هذا الخطأ خطراً ليهدد في حينها دقة التنبؤ الإحصائي لقيم المصفوفة الانتقالية؟ وما الشكل الذي ستؤول أليه المصفوفة p^n كما تم التطرق إلى تقدير المصفوفة $\lim_{n\to\infty} p^n$ ، وقد دعم البحث بجانبٍ تطبيقي تناول دراسةً حول منتجات الألبان المتاحة في الأسواق المحلية.

وبحسب معلومات الباحث فأن هذه الدراسة تعد الأولى في هذا المجال على المستوين المحلي والعالمي والتي تحاول دراسة هكذا نوع من القضايا.

١. *تمهيد (Introduction):*

يرفض البعض تماماً فكرة وجود الأخطاء ويعتقدون إن وجود الخطأ العشوائي أو الاضطراب (*Disturbance*) سببه الرئيسي هو ضعف الأداء في التجربة أو كما يعزيه البعض الآخر إلى ضعف أداء الباحث في طريقة تنفيذه للتجربة معتقدين أنه قد حدث في التجربة خطأً ما، في حين إن الإحصائي المتمرس يكون عادةً اكثر قدرةً على التعامل مع فكرة الأخطاء العشوائية من غيره ذلك إن أساس وجوده مرتبط بقوة بوجود حد الخطأ العشوائي وأن انعدام وجود هذا الحد سيؤدي بالضرورة إلى عدم شرعية وجود كلاً من الإحصاء والإحصائيين.

ويثار النقاش عادةً عن الجدوى المرجوة من دراسة الأخطاء وبنوعيها العشوائية وغير العشوائية إذ يتهيب البعض من دراسة الأخطاء ويعتقدون أن تقدير هذه الحدود قد يتطلب الكثير من المشقة والجهد الرياضي ليصل الباحث بعد ذلك إلى طرقٍ يتعذر عليه الولوج فيها مما يؤدي إلى توقف الباحث عن الاستمرار في بحثه، أن أصحاب هذه المخاوف والآراء لا يعون حقاً الجوهر الفعلي لأسلوب البحث العلمي والمعتمد على أسلوب التجربة والخطأ *(Try With Error)* ذلك لأنه ليس المطلوب دوماً إيجاد نتيجة أو حقيقة جديدة وإنما يكفي الباحث فخراً بأنه قد ولج طريقاً هو من طرق الشجعان يهاب الكثير الدخول فيه وأنه من خلال بحثه قد سجل أضافةً جديدةً للعلم وأن هذه الإضافة تتناسب مع كلاً من إمكاناته العلمية من جهة والتطور العلمي في ذلك الوقت من جهةٍ أخرى ،أن توقف الباحث عن الاستمرار في بحثه لا يعني بالضرورة فشل هذا الباحث بل على العكس من ذلك تماماً إذ يسجل له أنه قد أثار الطريق للآخرين للاستمرار من حيث توقف حتى أن ظهور نتائج غير واقعية يعد كذلك نصراً لهذا الباحث لأنه سوف يؤشر للآخرين بضرورة عدم الوقوع بنفس المطبات والمشاكل ،وللقارئ المتفحص في دقائق الأمور نجد إن الكثير من الباحثين عندما يلجون في كتابة بحوثهم وكتبهم فانهم يقومون بذلك على احسن وجه ولكن عند الاسترسال في موضوع الأخطاء في النماذج مضطرين فان أسلوب هذا الباحث يبدأ بالتخبط وعدم الانتظام في المسير الذي عهدناه له .

ومن هنا برزت فكرة البحث والذي يحاول إن يفسر فكرة الأخطاء غير العشوائية وبشكلٍ مبسط وبيان الفرق فيما بينها وبين الأخطاء العشوائية وسيتم التطرق بشكلٍ خاص إلى أخطاء التسجيل وشرح إمكانية تقديرها والتنبؤ بمسارها ومقدارها وذلك في واحدة من أهم أنواع التحاليل الإحصائية ألا وهو سلاسل ماركوف (*Markov Chain*) .

٢.*<u>هدف البحث (The Aim):</u>*

خصص هذا البحث لدراسة حالة وجود الأخطاء غير العشوائية في سلاسل ماركوف وأسباب حدوثها وبيان مدى تأثيرها على جميع عناصر المصفوفة الانتقالية (*Transition Matrix*) وذلك عند كلاً من المدى البعيد (*Long Run*) وحالة الثبات والاستقرارية (*Stationary*)، أي أن مجال اهتمام هذا البحث انصب على التنبؤ بمقدار الأخطاء غير العشوائية في مصفوفة الاحتمالات

الانتقالية عند الخطوة n والتركيز بشكلٍ خاص على أخطاء التسجيل لكونها من أصعب وأخطر أنواع الأخطاء غير العشوائية وهل سيشكل هذا الخطأ خطراً ليهدد في حينها دقة التنبؤ الإحصائي؟ وما الشكل الذي ستؤول أليه المصفوفة p^n ؟ كما تم التطرق إلى تقدير المصفوفة p^n وذلك عندما $n \to \infty$. ونظراً لمحدودية حيز العرض لذا قد تناول البحث في طياته دراسة حالة أخطاء التسجيل للمصفوفة الانتقالية من نوع 2 x 2 فقط مع جانبٍ تطبيقي تناول دراسة حول منتجات الألبان المتاحة في الأسواق المحلية وبحسب معلومات الباحث فأن هذه الدراسة تعد الأولى على كلاً من المستوين المحلي والعالمي والتي تحاول دراسة هكذا نوع من القضايا.

٣.*فلسفة الأخطاء العشوائية (The Logic of Random Errors) :*

يقسم كلاً من *Neter , Waeserman &Kunter* (1985) [10] نمط العلاقات *Relations* والتي تحدث في الطبيعة إلى قسمين رئيسين هما:

أ. العلاقات الرياضية (*Mathematical Function*) والتي تتصف بأن نمط العلاقة بين مدخلات الأنموذج ومخرجاته تكون مثالية أي انه عند معرفة مدخلات أي معادلة فان الحصول على مخرجاتها سيكون مباشراً وبصورة دقيقة لا تقبل الشك ومثالها

$$Y = a + b \qquad \text{.....(1)}$$

فعند معرفة قيمة a وb ولتكن على سبيل المثال 2 و3 فان قيمة Y لا بد أن تكون مساوية لـ 5 لذا فأن الأنموذج هنا يتصف بالمثالية.

ب. العلاقات الإحصائية (*Statistical Function*) والتي تتصف بأن نمط العلاقة فيها بين مدخلات الأنموذج ومخرجاته تكون غير مثالية أي أنه عند معرفة المدخلات للأنموذج فذلك لا يعني بالنتيجة التوصل إلى المخرجات بشكلٍ دقيق وإنما يجب اللجوء إلى إيجاد تقدير له باستخدام إحدى الطرق الإحصائية ومثالها

$$Y = a + b + \varepsilon \qquad \text{.....(2)}$$

حيث أن ε يمثل حد الخطأ العشوائي في الأنموذج.

وعند معرفة قيمة a وb والتي هي 2 و 3 فان قيمة Yستكون (في اغلب الحالات)لا تساوي 5 وإنما قد تكون اكبر منها أو اقل لذا فأن الأنموذج هنا لا يتصف بالمثالية.

أن الإضافة التي تحققت بإيجاد حد الخطأ العشوائي هي عظيمةٌ لدرجة لا يمكن وصفها فهو أساس النهضة العلمية الحديثة في العالم اليوم ذلك لان العلاقات الموجودة في الكون والطبيعة لا يمكن في أغلب الحالات للعلاقات الرياضية أن تفسر شرحاً أو وصفاً لها وذلك بسبب وجود حالات الاختلافات في الاستجابة في معظم الصفات والحالات البشرية و الطبيعية ...الخ. كما أن بعض العلاقات والتي يمكن أن توصف باستخدام نماذج رياضية (والتي هي محدودة بالطبع) ستكون معقدة جداً بحيث يستحيل عندها إيجاد حل للأنموذج وتفسير العلاقة من خلالها وذلك لأن حدود

الأنموذج ستكون أما بصيغة غير خطية معقدة أو أن تكون الحدود متداخلة لدرجة يصعب عندها حل الأنموذج، لذا فأننا نجد بان اكثر الحالات لا يمكن أن تشرح أو تفسر إلا من خلال استخدام العلاقات الإحصائية.

ومن هنا يمكنني أن أعرف حد الخطأ العشوائي ε على انه "ذلك الحد الذي سجل اضافةً كبيرة للعلم إذ انه وبواسطته استطاع العالم ومن خلال علمائه رصد جميع الحالات والظواهر والتي لم يكن في السابق من الممكن حسابها أو تقديرها بواسطة العلاقات الرياضية". وإجمالاً لكل ما ذكر فأننا نجد أن هنالك ترادفاً كبيراً بين مفهوم الأخطاء العشوائية من جهة والمعرفة الإحصائية من جهةٍ أخرى إذ إنني أعتقد بأنه أن كان ولابد من علامةٍ تجارية للإحصاء فلا يمكن أن تكون ألا حد الخطأ العشوائي (الابسيلون ε).

٤. *فلسفة الأخطاء غير العشوائية (The Logic of Non- Random Errors):*

وقد يتساءل شخص عن سبب حدوث الأخطاء غير العشوائية وهل هنالك فرق بينها وبين الأخطاء العشوائية ومن خلال الأدبيات والمصادر الإحصائية يمكن إجمال حالات حدوث الأخطاء غير العشوائية بالنقاط الآتية [6]:

١. أخطاء التسجيل. ٢.أخطاء القياس.

٣. أخطاء التجميع. ٤. الصيغة الناقصة في الأنموذج الرياضي.

إن الحالة الأولى والتي تمثل خطأً في التسجيل كانت قد طرحت في اكثر من موضع وحاولت الجمعيات والمؤسسات المختصة بالتطبيقات والمؤشرات الإحصائية جاهدةً أن تتصدى لهذا النوع من الأخطاء ،ذلك لان الإحصاء كما ورد تعريفه في قاموس أوكسفورد للكلمات على أنه "*حقائق تظهر بشكل أرقام تجمع وترتب لأجل المقارنة*" [8]، أي أن الإحصاء يستند أساساً إلى إيجاد مقدرات تصف معالم المجتمع باستخدام الأرقام المتاحة ولا يعتمد على نظريات الكتب ليكون بذلك حالةً فريدةً من نوعها ما بين العلوم المختلفة، إذ أنه الحالة الوحيدة والتي يستطيع الشخص من خلالها أن يعيش مع روح الرقم ليولد من خلاله معرفةً جديدةً لم تكن لتعرف من دونه[4].

أما عند استعراض الحالة الثانية من حالات حدوث الأخطاء العشوائية (أخطاء القياس) فيجب الوقوف ملياً على ما يذكره Huff[9] من أن التحليل الإحصائي الناجح يتكون من ثلاثة عناصر رئيسة هي:

١. إحصائي كفء يستطيع أن يفهم ما هو المطلوب فعلاً.

٢. اختيار التحليل الإحصائي الملائم للبيانات المتاحة.

٣. شخص يستطيع أن يفهم ويقدر الجدوى من التحليل الإحصائي والمؤشرات.

لذا فان خطأ القياس لا يمكن ابداً أن يتصاحب مع وجود إحصائي جيد لأن استخدام طريقةً ما في غير موضعها سيؤدي بالنتيجة إلى نتائج هزيلة تضعف التحليل الإحصائي ليس في دراسةٍ بحد

ذاتها وإنما في كل الإحصاء ذلك لأن هنالك من يتصيد الفرص للإحصاء والإحصائيين ليطلق العنان لأفكاره المريضة ويدعي بأن الإحصاء هو علم الكذابين والدجالين وكما في المقولة الشهيرة There (are three kinds of lies: lies, damned lies, and statistics)

وهنا من الواجب التأكيد على ضرورة أن يتأنى الإحصائيين في تحديد الطريقة الملائمة لتحليل البيانات المتاحة لهم إذ أنه ليس كل ما يتمناه المرء يناله[3].

أما الحالة الثالثة والخاصة بأخطاء التجميع فهذا أمرٌ قد عفا عليه الزمان إذ أن الاستبانات كانت تجمع وتفرغ سابقاً بطرقٍ قد يصاحبها التحيز مما يؤدي إلى عدم دقة البيانات التي يتم الحصول عليها ونجد في المصادر الحديثة بان العالم اليوم يميل إلى إيجاد حالة من الحصانة تبدأ أولاً بالتحديد الجيد لعينة الدراسة وبالشكل الذي تكون فيه ممثلة ليس عن المجتمع بالدرجة الأساس وإنما عن صفة ما في ذلك المجتمع والتي هي قيد الدراسة ومن ثم استخدام أفضل أنظمة المعلومات الحديثة في تفريغ البيانات [11] .

أما الحالة الرابعة والتي تمثل الصيغة الناقصة للأنموذج الرياضي فأننا نجد اليوم بان الإحصاء والذي أمسى باحثاً عن تخصصاتٍ دقيقة في مجال معرفته الضخمة قد عالج مثل هذه القضايا واوجد علمائه طرقاً لتحديد درجة دقة الأنموذج الرياضي الواجب استخدامه وفي جميع أصناف نظرياته وأساليبه من تحليل الانحدار وتصميم وتحليل التجارب ومتعدد المتغيرات ...الخ مما أدى إلى انحسار ظهور هكذا نوع من الأخطاء كذلك.

وإجمالاً لما ذكر فإننا نجد بان الحالة الأولى والتي تمثل أخطاء التسجيل تعد الحالة الأهم والأخطر وتكمن خطورة هذا النوع من الأخطاء العشوائية من كونه خطأً يصعب السيطرة عليه والحد من وجوده وأن وجود خطأً في التسجيل سيؤدي بالنتيجة إلى عدم كفاءة المقدرات أي أن وصف المجتمع سيكون متحيزاً وغير دقيق.

أما عند استعراض الفرق فيما بين فلسفتي الأخطاء العشوائية والأخطاء غير العشوائية من حيث المعالجة لكلاً منهما هو أن الأخطاء العشوائية هي أخطاءٌ لا يمكن التجاوز عنها وأن وجودها ضرورةً لا بد منها وهي عبارة عن الفروقات (التباينات والاختلافات) في الاستجابة ما بين المشاهدات أو الحالات في الدراسة وأن تعدد التحاليل الإحصائية يهدف بالدرجة الأساس إلى فظ الاشتباك بين المشاهدات وتفسير أكبر قدرٍ من الأخطاء العشوائية، أما الأخطاء غير العشوائية فهي أخطاءٌ لم يكن من الواجب ظهورها لذا كان من الواجب الحرص وبشكلٍ أكبر عند إجراء التحليل الإحصائي لمنع وجودها الأمر الذي يجعل الإحصائي يقف مكتوف اليدين أمام هكذا نوع من الأخطاء.

٥ .الجانب النظري (Theoretical Part):

يطلق مصطلح العملية التصادفية (Stochastic Process) أو العملية العشوائية (Random Process) أو عملية الفرصة (Chance Process) على أي ظاهرة حقيقية تجري أو تحدث في حيزٍ معين كالزمان أو المكان وذلك عندما تكون حالات ظهور تلك الظاهرة في أي وقت (أو لأي مشاهدة) تخضع لقوانين الاحتمالات[2]. وتعد العمليات التصادفية بشكلٍ عام وسلاسل ماركوف (Markov Chain) بشكلٍ خاص إحدى أهم الأساليب والتقنيات الإحصائية المستخدمة في التنبؤ (الاستشراق) وذلك من خلال إيجاد احتمالات البقاء في مرحلةٍ ما واحتمالات الانتقال بين المراحل المختلفة ولفتراتٍ زمنية مختلفة[1].

تقسم البيانات بشكلها العام في المعرفة الإحصائية إلى قسمين رئيسين القسم الأول هو البيانات المقطعية (Cross Section Data) والتي تهتم عادةً بدراسة صفة معينة يتصف بها المجتمع ويتطلب هذا النوع من البيانات أن يجمع من خلال أكثر من مصدر أو شخص بواسطة إحدى الطرق كالاستبانات ومثالها دراسة طريقة توزيع دخول مجموعة من الأفراد، أو إجراء مسح لحقوق ورغبات المستهلكين...الخ ،أما القسم الثاني من أنواع البيانات فهو بيانات السلاسل الزمنية (Time Series Data) والتي تهتم بدراسة تطور الحالة خلال الزمن لذا فأن هذا النوع من البيانات يجمع من مصدرٍ واحد ولفتراتٍ زمنيةٍ متتالية ومتساوية أي أن لها أبعاداً متساوية (سنوات، أشهر، أيام ...الخ) ومثالها دراسة إنتاج العراق من الحنطة للسنوات من سنة ١٩٨٠ ولغاية سنة ٢٠٠٢ واستخدام هذه البيانات للوصف والسيطرة خلال هذه المدة والاستشراق من خلالها لفتراتٍ زمنية لاحقة. وأن هذا التقسيم الرئيسي للبيانات في الإحصاء قد طال فيما طال سلاسل ماركوف كذلك، إذ أنها هي الأخرى تقسم بياناتها إلى قسمين رئيسين وفيما يلي تفصيلٌ لطرق احتساب قيم المصفوفة الانتقالية:

أ- سلاسل ماركوف ذات البيانات المقطعية:

تختلف سلاسل ماركوف في بنائها عن مجمل أصناف المعرفة الإحصائية بميزةٍ خاصة منحتها خصوصية ذلك بأن مصفوفة الاحتمالات الانتقالية تبنى على أساس حساب حالات الانتقال (التغير) من حالةٍ لأخرى وأن هذا التغير يتولد عبر الزمن أما بسبب ظرفٍ معين كتغير أسلوب الإنتاج، بث الإعلانات، حدثٌ اقتصادي معين ...الخ، أو من دون سبب كالرغبة في التغيير وهي الميزة الرئيسة في النفس البشرية ،وعند استخدام هذا النوع من البيانات في سلاسل ماركوف يتم اللجوء إلى استخدام وسيلتين للحصول على البيانات (استبانتين) حيث تقدم الاستبانة الأولى القراءة الأولى لكل شخص أو مفردة في مجتمع الدراسة وبعد مدةٍ زمنية معينة يحتسب التغير من خلال استبانه ثانية توزع لنفس الأشخاص الذين وزعت لهم الاستبانة الأولى وكما هو موضح في الجدول أدناه

الشخص	الاستبانة الأولى	الاستبانة الثانية

0	**1**	**1**
1	**1**	**2**
0	**0**	**3**
.	**.**	**.**
.	**.**	**.**
1	**0**	**M**

ب- سلاسل ماركوف ذات بيانات سلاسل زمنية :

وفي هذا النوع من سلاسل ماركوف يتم احتساب عناصر المصفوفة الانتقالية من خلال دراسة تطور الحالة عبر الزمن ولفترات زمنية متتالية وكما في دراسة تغير أسعار صرف العملات الأجنبية، تغير حجم وكم الإنتاج، عدد العاملين... الخ،وكما في الشكل الآتي

$$0 \longrightarrow 1 \longrightarrow 1 \longrightarrow 0 \longrightarrow 1 \longrightarrow \ldots\ldots\ldots\ldots \longrightarrow 1$$
$$t_1 \longrightarrow t_2 \longrightarrow t_3 \longrightarrow t_4 \longrightarrow t_5 \longrightarrow \ldots\ldots\ldots\ldots \longrightarrow t_m \qquad \ldots\ldots\ldots(3)$$

وسواء أكانت البيانات في سلاسل ماركوف بياناتٍ مقطعية أو سلاسل زمنية فأنه يشترط في المصفوفة الانتقالية أن تحقق الشروط الآتية:-

١. أن كل عنصر في المصفوفة الاحتمالية يشترط أن يكون قيمة احتمالية

$$0 \le p(\chi) \le 1 \qquad \ldots\ldots\ldots(4)$$

٢. أن مجموع كل صف من صفوف المصفوفة الانتقالية يساوي الواحد الصحيح أي أن كل صف هو p. m. f. *(Probability Mass Function)*.

$$P_{00} + P_{01} + P_{02} + P_{03} + \ldots\ldots + P_{0w} = 1$$
$$P_{10} + P_{11} + P_{12} + P_{13} + \ldots\ldots + P_{1w} = 1$$

.....(5)

$$P_{w0} + P_{w1} + P_{w2} + P_{w3} + \ldots\ldots + P_{ww} = 1$$

ونظراً لأن هذا البحث يصب مجال اهتماماته على سلاسل ماركوف ذات المرحلتين فقط *Two – states Markov chain* لذا فأن المصفوفة الانتقالية ستكون بالشكل الآتي:

$$p = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix} = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \qquad \ldots\ldots\ldots(6)$$

حيث أن $0 \le P_{00}, P_{01}, P_{10}, P_{11} \le 1$

وأن $P_{10} + P_{11} = 1$ **&** $P_{00} + P_{01} = 1$

الآن وبفرض حدوث أخطاءً في التسجيل في احتساب الانتقالات لعددٍ من المشاهدات أو الفترات الزمنية أدى إلى اختلال التوازن لقيمة واحدة من قيم المصفوفة الانتقالية الأربعة

لمصفوفة ماركوف ذات المرحلتين لذا فأن هنالك أربع حالاتٍ يحتمل فيها $(p_{00}, P_{01}, P_{10}, P_{11})$ ظهور هذا الخطأ وفيما يلي تفصيلٌ لها:-

أ- الحالة الأولى:

حدوث أخطاء في التسجيل في حساب مشاهدة أو اكثر من المشاهدات أدت إلى حصول زيادة في قيمة العنصر $1-a$ وبمقدار e وبذلك تكون المصفوفة الانتقالية بالشكل الآتي بدلاً عن شكلها الصحيح والمعروض في المعادلة(6):

$$\therefore P^* = \begin{pmatrix} 1-a+e & a-e \\ b & 1-b \end{pmatrix} \quad \text{.....(7)}$$

وعند احتساب المصفوفة الانتقالية عند الخطوة الثانية $\left(p^*\right)^2$ نجد أن قيم العناصر للمصفوفة قد أمست على النحو الآتي:

$$\begin{aligned} \therefore \left(P_{00}{}^*\right)^2 &= (1-a+e)(1-a+e)+(a-e)b \\ &= (1-a)^2 + ab + 2e - 2ae + e^2 - eb \end{aligned} \quad \text{.....(8)}$$

في حين أن قيمة $(p_{00})^2$ الفعلية يجب أن تكون مساوية لـ $(1-a)^2 + ab$ أي سجلت زيادة في قيمتها وقدرها $2e-2ae+e^2-eb$ ، وبنفس الطريقة لـ $\left(P_{01}{}^*\right)^2$

$$\begin{aligned} \left(P_{01}{}^*\right)^2 &= (a-e)(1-b)+(1-a+e)(a-e) \\ &= a[2-a-b]-2e+2ae-e^2+eb \end{aligned} \quad \text{.....(9)}$$

أي سجلت حالة نقصان في القيمة الفعلية $(p_{01})^2$ وقدرها $2e+2ae-e^2+eb$ ،ولـ $\left(P_{10}{}^*\right)^2$

$$\left(P_{10}{}^*\right)^2 = b[2-a-b]+eb \quad \text{.....(10)}$$

في حين أن القيمة الفعلية لـ $(p_{10})^2$ كان من المفترض أن تساوي $b[2-a-b]$ أي سجلت حالة زيادة في القيمة الحقيقية وقدرها eb . وبنفس الأسلوب نجد أنه أثناء تقدير $(p_{11})^2$ قد انخفضت قيمتها بمقدار eb ، وبوضع الاستنتاجات أعلاه في مصفوفة تتكون لدينا المصفوفة المعروضة في (11):

$$\left(P^*\right)^2 = \begin{pmatrix} P_{00}{}^2 + 2e(1-a)+e(1-b) & P_{01}{}^2 - 2e(1-a)-e(1-b) \\ P_{10}{}^2 + eb & P_{01}{}^2 - eb \end{pmatrix} \quad \text{.....(11)}$$

يتضح من المعادلة رقم (11) بأنه عند اختلال التوازن في قيمة واحدة من قيم المصفوفة الانتقالية فأن جميع عناصر المصفوفة عند الخطوة الثانية ستعاني هي الأخرى من حالة اختلال ولتعميم النتائج للحالة n بصورة عامة وكما هو معلوم فأن المصفوفة p^n يجب أن تكون[7]

$$P^n = \begin{pmatrix} \frac{b}{a+b} + \frac{a(1-a-b)^n}{a+b} & \frac{a}{a+b} - \frac{a(1-a-b)^n}{a+b} \\ \frac{b}{a+b} - \frac{b(1-a-b)^n}{a+b} & \frac{a}{a+b} + \frac{b(1-a-b)^n}{a+b} \end{pmatrix} \quad \text{.....(12)}$$

وفي حالة حصول زيادة في قيمة P_{00} وقدرها e ستؤول قيم المصفوفة $\left(P^*\right)^n$ إلى الشكل الآتي ✿:-

$$\left(P^*\right)^n = \begin{pmatrix} \frac{b}{a+b-e} + \frac{(a-e)(1-a-b+e)^n}{a+b-e} & \frac{a-e}{a+b-e} - \frac{(a-e)(1-a-b+e)^n}{a+b-e} \\ \frac{b}{a+b-e} - \frac{b(1-a-b+e)^n}{a+b-e} & \frac{a-e}{a+b-e} + \frac{b(1-a-b+e)^n}{a+b-e} \end{pmatrix}$$

.....(13)

وكما هو معلوم فعندما $n \to \infty$ تتحقق حالة الاستقرارية أو الثبات *Stationary* والتي يفترض أن تكون المصفوفة بالشكل الآتي✿✿:

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{pmatrix} \quad \text{.....(14)}$$

وفي الحالة الأولى تصبح مصفوفة الثبات بالشكل الآتي

$$\lim_{n\to\infty} p^n = \begin{pmatrix} \frac{b}{a+b-e} & \frac{a-e}{a+b-e} \\ \frac{b}{a+b-e} & \frac{a-e}{a+b-e} \end{pmatrix} \quad \text{.....(15)}$$

ومن هنا يتبين مدى الخطورة من وجود خطأً في تسجيل مشاهدة أو أكثر إذ حتى المصفوفة الانتقالية في المدى أو الخطوة n وذلك عندما تقترب nمن الما لانهاية سيظهر الخطأ وفي جميع القيم لعناصر المصفوفة الأمر الذي سيهدد دقة التنبؤ الإحصائي باستخدام سلاسل ماركوف.

<u>ب- الحالة الثانية:</u>

لتكن الحالة الثانية هو حصول زيادة في $P_{01} = a$ بمقدار e وبكلام آخر نقصان في P_{00} بنفس المقدار فستكون لدينا هنا عكس نتائج الحالة الأولى تماماً،وللاختصار سيتم عرض كلاً من المصفوفتين p^n والمصفوفة $\lim_{n\to\infty} p^n$ لكونها الأهم (والاشتقاق يكون بنفس الأسلوب المعروض في الملحقين الأول والثاني).

✿ **الاشتقاق كما هو مبين في الملحق رقم (١)**

✿✿ **الاشتقاق كما هو مبين في الملحق رقم (٢)**

$$\left(P^*\right)^n=\begin{pmatrix}\dfrac{b}{a+b+e}+\dfrac{(a+e)(1-a-b-e)^n}{a+b+e} & \dfrac{a+e}{a+b+e}-\dfrac{(a+e)(1-a-b-e)^n}{a+b+e}\\ \dfrac{b}{a+b+e}-\dfrac{b(1-a-b-e)^n}{a+b+e} & \dfrac{a+e}{a+b+e}+\dfrac{b(1-a-b-e)^n}{a+b+e}\end{pmatrix}$$

……..(16)

أما بالنسبة لحالة الثبات والاستقرارية $\lim_{n\to\infty} P^n$

$$\lim_{n\to\infty} P^n=\begin{pmatrix}\dfrac{b}{a+b+e} & \dfrac{a+e}{a+b+e}\\ \dfrac{b}{a+b+e} & \dfrac{a+e}{a+b+e}\end{pmatrix} \qquad \text{……(17)}$$

<u>ج – الحالة الثالثة:</u>

وبالانتقال إلى حالة حدوث خطأ في تسجيل أدى إلى زيادة قيمة P_{10} **بمقدار** e **لتصبح قيمته مساوية لـ** $P_{10}=b+e$ **ولتصبح المصفوفة بالشكل الآتي**

$$P^*=\begin{pmatrix}1-a & a\\ b+e & 1-b-e\end{pmatrix} \qquad \text{……(18)}$$

فأن المصفوفتين p^n **و** $\lim_{n\to\infty} P^n$ **ستكونان بالشكل الآتي:**

$$\left(P^*\right)^n=\begin{pmatrix}\dfrac{b+e}{a+b+e}+\dfrac{a}{a+b+e}(1-a-b-e)^n & \dfrac{a}{a+b+e}-\dfrac{a}{a+b+e}(1-a-b-e)^n\\ \dfrac{b+e}{a+b+e}-\dfrac{(b+e)}{a+b+e}(1-a-b-e)^n & \dfrac{a}{a+b+e}+\dfrac{(b+e)}{a+b+e}(1-a-b-e)^n\end{pmatrix}$$

……(19)

$$\lim_{n\to\infty} P^n=\begin{pmatrix}\dfrac{b+e}{a+b+e} & \dfrac{a}{a+b+e}\\ \dfrac{b+e}{a+b+e} & \dfrac{a}{a+b+e}\end{pmatrix} \qquad \text{……(20)}$$

<u>د – الحالة الرابعة:</u>

في حالة حصول زيادة في قيمة P_{11} **بمقدار** e **أي ستكون عكس الحالة الثالثة تماماً فأن المصفوفتين** p^n **و** $\lim_{n\to\infty} P^n$ **ستكونان بالشكل الآتي:**

$$\left(P^*\right)^n=\begin{pmatrix}\dfrac{b-e}{a+b-e}+\dfrac{a}{a+b-e}(1-a-b+e)^n & \dfrac{a}{a+b-e}-\dfrac{a}{a+b-e}(1-a-b+e)^n\\ \dfrac{b-e}{a+b-e}-\dfrac{(b-e)}{a+b-e}(1-a-b+e)^n & \dfrac{a}{a+b-e}+\dfrac{b-e}{a+b-e}(1-a-b+e)^n\end{pmatrix}$$

……(21)

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{b-e}{a+b-e} & \frac{a}{a+b-e} \\ \frac{b-e}{a+b-e} & \frac{a}{a+b-e} \end{pmatrix} \qquad \text{.....(22)}$$

٦ .الجانب التطبيقي (Applicable Part):

إن الميزة الرئيسة التي يمتاز بها الجانب التطبيقي لهذا البحث بأنه ليس جانباً أعد بالشكل الذي يخدم النظرية لكي يكملها بل على العكس من ذلك تماماً إذ أن فكرة هذا البحث قد تكونت وتطورت أثناء مشاركتي في إعداد أحد البحوث التطبيقية في مركز بحوث السوق وحماية المستهلك / جامعة بغداد ضمن خطته لسنة ٢٠٠٢ [5] ذلك انه خلال فترة إعداد البحث المذكور ظهرت مشكلة تطبيقية من واقع الحال تمثلت بأن عدداً من المشاهدات كانت قد سجلت وأدخلت بشكلٍ خاطئ الأمر الذي أدى ظهور خطأً في التسجيل ،ومن خلال مراجعة عدداً من المصادر والدوريات المتاحة لم أجد ما يدل على أخطاء التسجيل في سلاسل ماركوف لذا فقد عمدت إلى كتابة هذا البحث إذ كما هو معلوم "*لا خير في علمٍ لا ينتفع منه*".

هذا وقد تناول البحث المذكور استخدام سلاسل ماركوف في تحليل الأنماط السلوكية للمستهلك لمنتجات الألبان المتاحة في الأسواق المحلية [الاجبان بأنواعها، الحليب، القيمر (القشطة)] في محاولةٍ للكشف عن حالات التذبذب في سلوكية المستهلك العراقي الشرائية من حيث تفضيله لماركةٍ تجارية دون سواها والكشف عن حالات الانتقال في الاستخدام من نوعٍ لأخر (من ماركةٍ تجارية إلى ماركةٍ أخرى) ومن دون وجود سبب منطقي مبرر (بث الإعلانات، تغير الإنتاج، تغير الأسعار... الخ).حيث تم تحديد فترة زمنية وزعت استبانه على عينة من المستهلكين في مدينة بغداد وبعد مدةٍ زمنيةٍ معينة وزعت استبانه ثانية لنفس أفراد العينة من المستهلكين (بيانات مقطعية) ونظراً لأن البحث هو من ضمن البحوث المركزية والمعدة وفقاً لأهداف ومهام المركز لذا فقد تم اقتطاف جزءاً منه وبالشكل الذي لا يضر بالبحث وسرية البيانات ومن خلاله نجد انه قد تم احتساب مصفوفة الاحتمالات الانتقالية بالشكل التالي:

$$p^* = \begin{matrix} 0 \\ 1 \end{matrix}\begin{pmatrix} 0.65 & 0.35 \\ 0.70 & 0.3 \end{pmatrix}$$

إذ أن 0 ترمز لنوعية القيمر (القشطة) الأكثر استخداما (الشركة ذات الموقع التنافسي الأقوى)

1 ترمز لباقي أنواع القيمر المتاحة في الأسواق المحلية.

في حين أن مصفوفة البيانات الصحيحة والتي هي خالية من أخطاء التسجيل يجب أن تكون بالشكل الآتي :

$$p = \begin{matrix}0\\1\end{matrix}\begin{pmatrix}0.8 & 0.2\\0.7 & 0.3\end{pmatrix}$$

مما يعني وجود خطأ في التسجيل من الحالة الثانية ومقداره $e = 0.15$ ،والجدول الآتي يبين تقدير كلاً من المصفوفة الانتقالية الصحيحة (الخالية من أخطاء التسجيل) والمصفوفة التي تحوي أخطاءً في التسجيل وذلك وفقاً للخطوات:

Step	P	P^*
1	$\begin{pmatrix}0.8 & 0.2\\0.7 & 0.3\end{pmatrix}$	$\begin{pmatrix}0.65 & 0.35\\0.70 & 0.3\end{pmatrix}$
2	$\begin{pmatrix}0.78 & 0.22\\0.77 & 0.23\end{pmatrix}$	$\begin{pmatrix}0.6675 & 0.3325\\0.665 & 0.335\end{pmatrix}$
3	$\begin{pmatrix}0.778 & 0.222\\0.777 & 0.223\end{pmatrix}$	$\begin{pmatrix}0.666625 & 0.333375\\0.66675 & 0.33325\end{pmatrix}$
4	$\begin{pmatrix}0.7778 & 0.2222\\0.7777 & 0.2223\end{pmatrix}$	$\begin{pmatrix}0.66666875 & 0.33333125\\0.6666625 & 0.3333375\end{pmatrix}$
$\vdots$	$\vdots$	$\vdots$
n	$\begin{pmatrix}\frac{0.7+0.2(0.1)^n}{0.9} & \frac{0.2-0.2(0.1)^n}{0.9}\\\frac{0.7-0.7(0.1)^n}{0.9} & \frac{0.2+0.7(0.1)^n}{0.9}\end{pmatrix}$	$\begin{pmatrix}\frac{0.7+0.35(-0.05)^n}{1.05} & \frac{0.35-0.35(-0.05)^n}{1.05}\\\frac{0.7-0.7(-0.05)^n}{1.05} & \frac{0.35+0.7(-0.05)^n}{1.05}\end{pmatrix}$
$\lim_{n\to\infty}$	$\begin{pmatrix}0.777777777778 & 0.222222222222\\0.777777777778 & 0.222222222222\end{pmatrix}$	$\begin{pmatrix}0.666666666667 & 0.333333333333\\0.666666666667 & 0.333333333333\end{pmatrix}$

ومن الجدول أعلاه نلحظ مدى خطورة وجود أخطاء في التسجيل وقد تبين تأثيرها على دقة وكفاءة التنبؤ الإحصائي باستخدام سلاسل ماركوف ،ومن الجدير بالذكر هنا أن القيم لعناصر المصفوفات في أعلاه قد طابقت تماماً الصيغ والمعادلات الرياضية المشتقة مما يؤكد على دقتها وصحتها.

٧ .الاستنتاجات والتوصيات:

أ . الاستنتاجات:

١. يبين هذا البحث مدى خطورة وجود أخطاء التسجيل عند استخدام سلاسل ماركوف وذلك لأن وجود خطأً واحداً فقط في احتساب حالات الانتقال من مشاهدةٍ لأخرى أو من فترة زمنية لأخرى سيؤدي بالنتيجة إلى تدوير الخطأ وانتشاره لجميع قيم المصفوفة وذلك عند p^2 الأمر الذي حثني على تصور الحالة عند الخطوة (p^n) وكان الأمر مشابهاً تماماً لما تم تصوره فعلاً حول خطورة الموقف حتى أنه عندما تقترب n من المالا نهاية $(n \to \infty)$فأن قيمة الخطأ لم تضمحل كما كنت أعتقد على أقل تقدير.

٢. تبين من خلال هذا البحث بأنه يستحيل القضاء على حد الخطأ العشوائي وأن انتشاره سيعني بالضرورة إلى أضعاف الثقة بالتقديرات والتنبؤات الناجمة عن استخدام سلاسل ماركوف وأن هذا الابتعاد سيكون معتمداً بالدرجة الأساس على مقدار حد الخطأ e وبالاستفادة من الخاصية الاحتمالية لسلاسل ماركوف ذلك أنها لا تستخدم الكميات بل الاحتمالات كقيم لعناصر المصفوفة الانتقالية لذا فأن زيادة حجم العينة (أما عدد المشاهدات أو طول الفترة الزمنية) سيؤدي بالنتيجة إلى تقليل وتهميش حد الخطأ (تقليل نسبة وجوده) الأمر الذي يقودنا إلى رفع الجودة للمقدرات وزيادة درجة الثقة بها.

٣. أن احتمال ظهور أخطاء التسجيل في سلاسل ماركوف هو أمرُ دارج ومحتمل بشكلٍ كبير خصوصاً وأن جميع البرمجيات الجاهزة باستخدام الحاسوب تكاد تخلو من ذكر سلاسل ماركوف مما يعني القيام باحتساب مصفوفة الانتقالات p يدوياً الأمر الذي يؤدي إلى ظهور أخطاء التجميع كذلك .

ب . التوصيات:

١. ضرورة التأكيد على المساحين والباحثين بزيادة الحرص عند تسجيل المشاهدات وذلك في محاولةٍ للحد من احتمال ظهور خطأً في التسجيل وخصوصاً عند استخدام سلاسل ماركوف.

٢. تناول البحث في طياته دراسة احتمال حدوث خطأ في التسجيل أدى إلى اختلال المصفوفة الانتقالية ذات المرحلتين في أحد صفوفها لذا وكدراساتٍ لاحقة لهذا البحث أوصي الباحثين بالاستمرار بالدراسة لاحتمال حدوث اكثر من خطأ في المصفوفة الانتقالية

$$P^* = \begin{pmatrix} 1-a+e & a-e \\ b+e' & 1-b-e' \end{pmatrix}$$

إذ أن e يمثل الخطأ الأول .

وأن e' يمثل الخطأ الثاني .

كما يمكن تناول أخطاء التسجيل في سلاسل ماركوف ذات ثلاثة مراحل فالمراحل الأكبر مدعمةً بجانب تجريبي باستخدام المحاكاة وذلك لوضع صيغة عامة يمكن تعميمها على جميع أنواع السلاسل .

٣.أختص هذا البحث بدراسة أخطاء التسجيل في سلاسل ماركوف ولكن بأجراء بعض التعديلات يمكن للباحثين الاستفادة من الأسلوب (بعد أجراء بعض التعديلات) لمعالجة كلاً من مشكلتي البيانات غير التامة (*Incomplete Data Problem*)والبيانات الشاذة (*Outliers*) في سلاسل ماركوف.

٤. لأغراض السهولة يعمد البعض إلى استخدام عملية تقريب الأرقام في سلاسل ماركوف وأوصي هنا بضرورة الابتعاد عن أي تقريب للأرقام وخصوصاً عند استخدام سلاسل ماركوف لأن خطأً غير عشوائي سيظهر هنا وسيدور هذا الخطأ للمراحل والخطوات اللاحقة الأمر الذي سيؤدي إلى أضعاف دقة التنبؤ الإحصائي .

المصادر (References):

١. البازي، عمار فردريك والرواف، محمد عبد الكريم (٢٠٠٢) " تصميم أنموذج ماركوفي خزني لحركة المياه" مجلة المنصور /المجلد الأول _ العدد الخامس.

٢. الربيعي، د. فاضل محسن وعبد، د. صلاح حمزة (٢٠٠٠) "مقدمة في العمليات التصادفية"، مديرية دار الكتب للطباعة والنشر، بغداد.

٣. الربيعي، د. ليث سلمان والبياع، مهدي محمد (٢٠٠١) "أخطاء بعض المعالجات الإحصائية للبيانات اللامعلمية في العلوم الإدارية"، مجلة العلوم الاقتصادية والإدارية، جامعة بغداد، عدد خاص بوقائع المؤتمر القطري الأول للعلوم الإدارية.

٤. الناصر، د. عبد المجيد حمزة والبياع، مهدي محمد (٢٠٠٢) "الإحصاء معرفة / سلسلة بحوث رقم (١)"، المؤتمر القطري الثالث عشر للجمعية العراقية للعلوم الإحصاء.

٥. خلف، د.عبد الرحيم والوائلي، د. منعم عزيز والبياع، مهدي محمد وشاهر، ثائر فيصل (٢٠٠٢) "استخدام سلاسل ماركوف بتحليل سلوك المستهلك (دراسة تطبيقية)" دراسة منجزة في مركز بحوث السوق وحماية المستهلك ضمن خطة عام ٢٠٠٢.

٦. كاظم، د.أموري هادي وعيفان، محمد مناجد (١٩٨٨) " مقدمة في تحليل الانحدار الخطي" مطبعة جامعة بغداد.

7. Bhat, f. (1972) "Elements of Applied Stochastic Processes" John Wiley & Sons.

8. Hornby, A.S. &Gatenby, E.V. & Wakefield, H (1963) "The Advanced Learner's Dictionary of Current English" Oxford University.

9. Huff, D. (1954) "How to lie with statistics" Victor Gollancz Limited.

10. Neter, J & Wasserman, W. & Kunter, M. H. (1985) "Applied linear statistical models" 2nd Edition, IRWIN, USA.

11. Wilcox, R. R. (1997)"Introduction to Robust Estimation and Hypothesis Testing", Academic press.

الملحق (Appendix):

Appendix (1)

$$\because P_{00}^{(n)} = P_{00}^{(n-1)}P_{00} + P_{01}^{(n-1)}P_{10}$$

$$= (1-a+e)P_{00}^{(n-1)} + b\left(1 - P_{00}^{(n-1)}\right)$$

$$\therefore P_{00}^{(n)} = b + (1-a-b+e)P_{00}^{(n-1)}$$

$$P_{00}^{(n)} = b + b(1-a-b+e) + b(1-a-b+e)^2 + \ldots.. (1-a+e)(1-a-b+e)^{(n-1)}$$

$$= b\sum_{r=0}^{n-2}(1-a-b+e)^r + (1-a+e)(1-a-b+e)^{n-1}$$

$$\because \sum_{r=0}^{n-2}(1-a-b+e)^r = \frac{1-(1-a-b+e)^{n-1}}{1-(1-a-b+e)}$$ **[قانون المتوالية الهندسية]**

$$\therefore P_{00}{}^{n} = \frac{b}{a+b-e} + \frac{(1-a-b+e)^{n-1}\left[a - a^2 - ab + ae - e + ae + be - e^2\right]}{a+b-e}$$

$$= \frac{b}{a+b-e} + \frac{a(1-a-b+e)^n - e(1-a-b+e)^n}{a+b-e}$$

$$\therefore P_{00}{}^{n} = \frac{b}{a+b-e} + \frac{(a-e)(1-a-b+e)^n}{a+b-e}$$

$$\because P_{00}{}^{n} + P_{01}{}^{n} = 1$$

$$\therefore P_{01}{}^{n} = \frac{a-e}{a+b-e} + \frac{(a-e)(1-a-b+e)^n}{a+b-e}$$

وبنفس الأسلوب يتم حساب $P_{10}{}^{n}$ ومن خلالها حساب $P_{11}{}^{n}$.

Appendix (2)

$$\lim_{n\to\infty} p_{00}{}^{*(n)} = \lim_{n\to\infty} p_{00}{}^{*(n-1)} = \prod_{00}$$

$$\because P_{00}{}^{*(n)} = b + (1-a-b+e)P_{00}{}^{*(n-1)}$$

$$\therefore \prod_{00}{}^{(n)} = b + (1-a-b+e)\prod_{00}{}^{(n-1)}$$

$$\therefore \prod_{00} = \frac{b}{a+b-e}$$

$$\because \prod_{01} = 1 - \prod_{00}$$

$$\therefore \prod_{01} = \frac{a-e}{a+b-e}$$

وبنفس الأسلوب تماماً نجد كلاً من $\lim_{n\to\infty} p_{10}^{*(n)}$ **و** $\lim_{n\to\infty} p_{11}^{*(n)}$.

Mistakes of the registration are in Markov Chain \an applied study

By

Al-Bayaa, Mahdi M.

The Market research and consumer Protection center
Baghdad University

Summary

The idea of the mistakes becomes by form the exciting thinks are for the discussion whether it charging with the ascription is statisticians or the researchers have who many is the scopes and who trade with one another with this idea where believes all by that the subject of the mistakes it is from the difficult and complex thesis, and from here an idea came out this discussion and which deceives by pretenses come you explain the idea of the mistakes non random and simplified is by form and a statement of the differentiation between it and between the mistakes random to one from the most dangerous of the kinds of the mistakes non random made every effort and it is mistakes of the registration its appreciation cut and prediction in slices by its path and than my one from important kinds of the statistical analysis that is ((Markov Chain)) and the statement of an effect extent is the presence of the registration mistakes on the values of the Transition matrix elements and that is on the long run and a situation of the firmness and Stationary .In other words that the field of an interest is this discussion raise prediction highness amounting to the mistakes are non random in the matrix of the eventualities Transition the recalcitrant of the step n and does this error fashion a danger to crash down when as its prediction accuracy my counting for the values of the Transition matrix and what is the form which will return to the matrix p^n ? And then we study the estimation of the matrix $\lim_{n\to\infty} p^n$, and it had supported the discussion with an application study about the products of the available milks in the markets.

The researcher is according to an information then that this study becomes the first on both local and world levels.