المجلة العراقية للعلوم الإحصائية (27) 2018
ص ص [107-120]

استخدام سلاسل ماركوف في المجالات الطبية

عبدالغفور جاسم سالم العبيدي* عمار ياسين سليمان**
Tcsm_college@yahoo.com Tcsm_college@yahoo.com

المستخلص :

تم في هذا البحث دراسة السلسلة الزمنية لعدد الإصابات بمرض ذات الرئة كمتسلسلة ماركوف ، وذلك بوضع افتراضات على عدد الإصابات لصياغة المسألة وفق نموذج متسلسلة ماركوف بالاعتماد على عدد الحالات التي تمثل الظاهر. وبعد إيجاد الصفات الإحصائية لهذه السلسلة تبين أنها ثبوتية (Ergodic) ، وتم إيجاد التوزيع المستقر (Stationary distribution) لهذه السلسلة .

الكلمات الدلالية : سلسلة ماركوف، الثبوتية ، التوزيع المستقر .

Using Markov Chains in Medical Field

Abstract:

This paper studies the time series for a number of pneumonia patient as a Markov chain by proposing hypotheses concerning the number of infections in order to set the problem according to Markov Chain depending on the number of cases that are represented. After we found the statistical characteristics for the chain that proved to be (Ergodic), then we found the stationary distribution for it.

*استاذ / قسم الرياضيات/ كلية علوم الحاسوب والرياضيات / جامعة الموصل
*بكلوريوس / قسم الرياضيات/ كلية علوم الحاسوب والرياضيات / جامعة الموص

تاريخ استلام البحث 2018/6/7 تاريخ القبول 2018/7/3

1-: المقدمة:

إنَّ متسلسلة ماركوف (Markov chain) هي حالة خاصة من العملية التصادفية (العشوائية) (Stochastic process) التي تشمل عدداً من الحالات (States) ، ويعد العالم Markov (1907) من الرواد في وضع المفاهيم الأساسية لمتسلسلة ماركوف ، طور هذه المفاهيم العديد من الرياضيين .

واستخدمت متسلسلة ماركوف في شتى المجالات (الزراعية ، والصناعية ، والتجارية ،..) تركز اهتمامنا في هذا البحث على صياغة عدد الإصابات بمرض ذات الرئة في مدينة الموصل كمتسلسلة ماركوف ، وتم إيجاد بعض الصفات الإحصائية لهذه السلسلة .

هناك الكثير من العمليات التصادفية (Stochastic process) التي نمر بها في الحياة اليومية من بينها العمليات المسماة عمليات ماركوف ، التي لها مكانة هامة في التطبيقات الاحصائية . وعمم مفهوم العملية التصادفية (Stochastic Process) لكي يشمل أية ظاهرة يتغير حدوثها بتغير الزمن (سواء كان حتمياً أم احتمالياً أم جوهرياً) والقابل للتحليل من ناحية الاحتمالية . وتسمى أحياناً عملية عشوائية (Random Process) ، وهي النظير الى العملية الحتمية (أو النظام الحتمي) فبدلاً من التعامل مع حقيقة واحدة محتملة في كيفية تطور هكذا عملية بتغير الزمن فإنه في العملية التصادفية أو العملية العشوائية يكون هناك جانب غير محدد في تطوره المستقبلي . إن سلسلة ماركوف (Markov chain) هي حالة خاصة من العملية التصادفية (Stochastic Process)، التي تحتوي على عدد محدود أو غير محدود من الحالات (State). وعلى هذا الأساس يمكن تحويل أية عملية تصادفية إلى عملية ماركوف ، وذلك بتغير مفهوم الحالة بما يتناسب مع السلسلة الزمنية المعطاة .

يعد العالم الروسي A.A Markov (1906–1970) من الرواد في وضع المفاهيم الاساسية لسلسلة ماركوف [2] . ثم طورت على يد العديد من الرياضيين منهم . الباحث (N. Wiener)

إذ أدخل في عام 1923 الهيكل الرياضي الصحيح لعملية ماركوف ذات المسارات المتصلة ، وأطلق على هذه العملية عملية (Wiener Process) [2] .

ثم توالت البحوث والدراسات وبشتى المجالات وباستخدام متسلسلات ماركوف .

وتناول الباحثان (Kaldfleisch ,and Lawless 1985) تحليل البيانات المجدولة على ضوء افتراض ماركوف مستمرة الزمن ، مثل مقدار دالة الإمكان الأعظم ومصفوفة التباين المكافئة لها[2] .

وفي 1988 قام الباحثان (Tavarea and Souza) بدراسة الجفاف في مدينة (Fartaleza) شمال غرب البرازيل بوصف سلسلة الأمطار كمتسلسلة ماركوف ذات الحالتين (الرطبة و الجافة) [8] .

وفي العام 1996 قام الباحثان (Salim and Thanoon) بدراسة نموذج سلاسل ماركوف لمنسوب نهر بثلاث حالات.

وفي العام 2005 قام العبيدي ، عبد الغفور جاسم وآخرون بدراسة الهجرة السكانية بين محافظات العراق الثلاث (نينوى وصلاح الدين والتأميم) . وهناك دراسات عديدة في هذا المجال. وسنحاول في هذا البحث دراسة السلسلة الزمنية لعدد الإصابات بمرض ذات الرئة وصياغتها وفق نموذج ماركوف.

2- : الهدف من البحث

يهدف البحث إلى دراسة متسلسلة ماركوف المتمثلة بعدد الإصابات بمرض ذات الرئة في مدينة الموصل (كمعدلات شهرية)، وصياغة المسألة وفق نموذج ماركوفي، ودراسة صفات المتسلسلة الناتجة من ذلك .

3-: تعاريف ومبادئ أساسية :

(1): العملية التصادفية (العشوائية) (Stochastic Process) :

هي مجموعة (عائلة) من المتغيرات العشوائية التي يستدل بها بالدليل t ، إذ إن t يعود الى مجموعة دليليه T (index set) ويرمز لذلك عادة بالرمز $\{X(t),\ t\in T\}$ أو للاختصار $\{X(t)\}$. [2]

(2) المتسلسلة الزمنية (Time series) :

إذا كانت المجموعة الدليلية T في العملية التصادفية (Stochastic Process) تمثل الزمن فإن العملية العشوائية تسمى عندئذ بالسلسلة الزمنية (Time series) ، وتكون السلسلة الزمنية مستمرة الزمن (Continuous time) إذا كانت $-\infty < t < \infty$ ، ويرمز لها$\{X(t)\}$ واذا كانت t قيماً متقطعة $t = 0,\pm1,\pm2,...$ فإن السلسلة الزمنية تسمى عندئذ سلسلة متقطعة الزمن (Discrete time series) ، ويرمز لها الرمز $\{X_t, t = 0,\pm1,\pm2,...\}$ أو للاختصار بـ $\{X_t\}$ ، وتكون السلسلة مستمرة إذا كانت العزوم الاحصائية لا تتأثر بالزمن ، كما يكون للسلسلة اتجاه عام إذا كانت تتذبذب حول محور وهمي لا يوازي محور السينات [2] .

(3) عملية ماركوف (Markov process) :

تسمى العملية العشوائية (التصادفية) ذات المعلمة التي تدل على الزمن المتقطع أو الزمن المستمر بعملية ماركوف (Markove Process) إذا حققت خاصية ماركوف (Markov Property) وهو الشرط الآتي:

إذا كان التوزيع الاحتمالي الشرطي لـ X_{tn} ، ولأية مجموعة معطاة من القيم $x_{t1}, x_{t2}, \dots x_{tn-1}$ فيعتمد فقط على x_{tn-1} ، ولأية مجموعة من الفترات الزمنية $t_1 < t_2 < \cdots < t_n$ وبمعنى أدق :

$$\mathrm{P}\left[X_{tn} = X_n / X_{t1} = X_{1,} X_{t2,} = X_2, \dots, X_{tn-1} = X_{n-1}\right] = \mathrm{P}[X_{tn} = X_n / X_{tn-1} = X_{n-1}]$$

لأي عدد من الأعداد الحقيقية X_1 , X_2......., X_n . أي إن الحالة المستقبلية تكون مستقلة عن الحالات السابقة بشرط أن تكون الحالة الحالية X_{n-1} معروفة [2] و [3] .

(4) متسلسلة ماركوف (Markov chain) :

متسلسلة ماركوف (Markov chain) هي حالة خاصة من العملية التصادفية (Stochastic Process) ذات عدد محدد ، أو غير محدود من الحالات (State) وان المعلمة t (Parameter) التي تميزها تدل على الزمن .

ويمكن عَدُّ سلسلة ماركوف هي سلسلة من الحالات التي تمر بها ظاهرة ما خلال مدة زمنية معينة. أو سلسلة من المواقع التي يمر بها جسم متحرك خلال مدد زمنية مختلفة [2] و [4].

وبناء على ذلك ستكون لدينا حالتان :

الحالة الأولى : هي حالة الزمن المتقطع (Discrete time Markova chain) ، وفي هذه الحالة هناك نظام معين يلاحظ في فترات زمنية متقطعة ، وإن مجموعة المشاهدات يمكن تمثيلها ما يأتي, X_n X_o, $X_{1,\ldots,}$ بافتراض أَنَّ X_{tn} تمثل متغيراً عشوائياً تدل قيمته على الحالة عند الزمن (n) فتسلسل المتغيرات X_{tn} يسمى بمتسلسلة ماركوف .

الحالة الثانية : هي حالة كون المعلمة تدل على مدد الزمن المستمر (Continuous time markov chain) ، إذ يفترض بأَنَّ المعلمة التي تدل على الزمن عبارة عن مجموعة من الارقام الحقيقية غير السالبة .

(5) المصفوفة البدائية (Primitive Matrix):

تسمى مصفوفة احتمالات الانتقال P بالمصفوفة البدائية (Primitive Matrix) إذا احتوت على قيمة ذاتية (Eigen Value) التي تكون مساوية للواحد ، و أَن بقية القيم الذاتية الأخرى أقل منها ، أي أَن $\lambda_1 > \lambda_i$ ، i=2,3,.... بالقيمة المطلقة. [5] ويمكن إيجاد القيم الذاتية من حل النظام

$$|P - \lambda I| = 0 \quad \ldots (5\text{-}1)$$

(6) المصفوفة غير القابلة للتجزئة (Irreducible matrix) :

تكون مصفوفة الانتقال قابلة للتجزئة (reducible) إذا كان بالإمكان إيجاد مصفوفة جزئية مغلقة من هذه المصفوفة ، وبعكسه تكون المصفوفة الانتقالية غير قابلة للتجزئة (Irreducible matrix) أي إن :

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \qquad \ldots \qquad (6\text{-}1)$$

فإذا كانت A_{11} , A_{22} مجموعات مغلقة فإنَّ A تكون قابلة للتجزئة (Reducible) ، وبعكسه $A=A_{11}$ أي أَن A هي المجموعة المغلقة الوحيدة . [6]

(7) المجموعة المغلقة (Closed set)

لتكن C مجموعة جزئية من فضاء الحالة S . تكون C مغلقة (Closed) إذا كان $P^{(n)}{}_{ik} = 0$ إذ $i \in C, k \neq 0$.لكل قيم n الصحيحة الموجبة [4] .

8) حالة المبادلة والاتصال (Communication) :

يقال للحالتين i و j أَنهما قابلتان للاتصال (Communication) ، ويرمز لها بالرمز $i \leftrightarrow j$ إذا وجد عددان صحيحان $n, m \geq 1$ بحيث أن $P^{(n)}{}_{ij} > 0$ و $P^{(m)}{}_{ji} > 0$ أي إنه يمكن للسلسلة الانتقال من الحالة i إلى الحالة j بعد n من الخطوات ، وكذلك بإمكانها الانتقال من الحالة j الى الحالة i بعد m من الخطوات و n ,m ليس بالضرورة ان يكونا متساويين [4] .

(9) مبرهنة (1):

تكون متسلسلة ماركوف غير قابلة للتجزئة إذا وفقط إذا كانت جميع الحالات متصلة . [6] .

(10) : السلسلة الثبوتية (Ergodic chain) :

تكون سلسلة ماركوف ثبوتية (Ergodic) اذا كان من الممكن الانتقال من كل حالة الى كل حالة (ليس بالضرورة ان يكون الانتقال بخطوة واحدة) ، أو بعبارة أخرى سلسلة ماركوف تكون ثبوتية (Ergodic) إذا كانت بدائية (Primitive) وغير قابلة للتجزئة (Irreducible). [3] .

(11) : التوزيع المستقر (Stationary distribution) : [3]

إذا كانت سلسلة ماركوف (Markov chain) غير قابلة للتجزئة (Irreducible) ،وتحتوي على حالات عودة (recurrent state) فإنها تمتلك توزيعاً مستقراً π ووحيداً . ويمكن إيجاده بحل النظام الآتي :

أي إن :

$$\left.\begin{aligned} &\pi = \pi P \\ &\sum_{j=1}^{n} \pi_j = 1 \end{aligned}\right\} \qquad ....(11-1)$$

$$\pi_j = \lim P^{n}{}_{ij} > 0,\ j = 0,1,2,... \qquad ...(11-2)$$

4:-الجانب الإحصائي والتطبيقي :

نحاول في هذه الفقرة دراسة السلسلة الزمنية لعدد الإصابات بمرض ذات الرئة في مدينة الموصل وفق نموذج متسلسلة ماركوف ، فقد تم الحصول على البيانات من (مستشفى السلام التعليمي في محافظة نينوى) المتمثلة بالمعدلات الشهرية لأعداد المصابين بالمرض (ذكوراً وإناثاً) للمدة (2009−2012) ، وبمختلف الفئات العمرية وسنحاول تحويل البيانات إلى صيغة متسلسلة ماركوف وكما يأتي :

الخطوة الأولى للمسألة هي تعريف الخطوة (Step) والحالة (State) لعدد المصابين لتستخدم في متسلسلة ماركوف . ثم وضع عدة افتراضات لتشمل الحالات (States) ، وكيفية حركة هذه الحالات التي تمثل أعداد المرضى لتكوين مصفوفة الانتقال (Transition matrix) .

وتم في هذا البحث تعريف الحالة (State) بأنها عدد المصابين بهذا المرض لمدة أربع سنوات حسب البيانات المتوفرة .

أما الخطوة (Step) فتمثل زيادة عدد المصابين من مدة زمنية إلى أخرى .

والجدول (1) يمثل فئات العمر المختلفة التي تتراوح بين سنة واحدة إلى أكثر من 65 سنة

جدول (1): يبين الفئات العمرية

Class	State
14-0	S_1
44-15	S_2
64-45	S_3
and more 65	S_4

وبالاعتماد على هذه الافتراضات تم الحصول على الجدول الآتي الذي يمثل عدد المرضى (ذكوراً وإناثاً) المصابين بمرض ذات الرئة بعد تبويبها .

جدول (2):يبين اعداد المرضى موزعين على الحالات (الفئات العمرية)

S_4	S_3	S_2	S_1	
8	27	54	60	2009
23	43	75	143	2010
47	60	187	206	2011
36	42	123	201	2012

إذ إن $S_i , i = 1,2,3,4$ تمثل الحالات المختلفة لمتسلسلة ماركوف (فئات العمر) .

تكوين مصفوفة الانتقال :

لتكوين المصفوفة الاحتمالية: (Transition probability matrix) ، التي هي مصفوفة الانتقال من أية حالة إلى أخرى في وحدة زمنية واحدة .

نحتاج إلى بيانات تصف حركة المرضى بصورة منفردة عبر الزمن ، وهذه البيانات غير متوفرة لدينا لسبب عدم مراجعة المرضى بشكل دوري للمراكز الصحية .

أما المتوافر من البيانات فيعطي معلومات عن العدد الكلي للمرضى بفئات العمر المختلفة ، واتخذت المدة الزمنية (سنة واحدة) كمدة أساسية ملائمة لمصفوفة التحول التي يمكن تكوينها بوضع افتراضات ملائمة لوصف حركة المرضى بين الفئات العمرية المختلفة وهي :

- أي مريض يصل إلى الحالة S4 يبقى ضمنها .
- أي مريض يزداد عمره سينتقل إلى المستوى الذي يكون أعلى منه ، فالزيادة في عدد المرضى في أية حالة S_i تأتي من الحالة السابقة لها مباشرة S_{i-1} .
- التناقص في عدد المرضى ناتج عن انتقالهم إلى الحالة S_0 المتمثلة بحالة الشفاء.

وباستخدام هذه الافتراضات والبيانات المتوفرة لمدة سنة واحدة (مدة الأساس) تم الحصول على الجدول (3) الذي يمثل أعداد المرضى ضمن الحالات للسنوات (2010−2009) .

الجدول (3) يبين أعداد المرض ضمن الحالات للسنوات (2010−2009)

States	S_1	S_2	S_3	S_4
2009	60	54	27	8
2010	143	75	43	23

لذا يمكننا تبيان الحركة التقديرية للمرضى من حالة إلى أخرى خلال (2010−2009) كما هي موضحة في الجدول (4) .

جدول (4) يبين الحركة التقديرية للمرضى للمدة (2009–2010)

2009/2010	S_0	S_1	S_2	S_3	S_4	الصفوف
S_0	0	0	0	0	0	0
S_1	135	8	0	0	0	143
S_2	0	52	23	0	0	75
S_3	0	0	31	12	0	43
S_4	0	0	0	15	8	23
الأعمدة	135	60	54	27	8	

وبالطريقة نفسها يتم تكوين الجداول الأخرى للسنوات 2011–2010 و 2012–2011 .

وبجمع الجداول وفقاً لقاعدة جمع المصفوفات نحصل على الجدول الآتي (الجدول 5) الذي يوضح التحولات (الحركات) التقديرية للمرضى من عام (2009) لغاية (2012) .

الجدول (5) يبين الحركات التقديرية للمرضى (2012–2009)

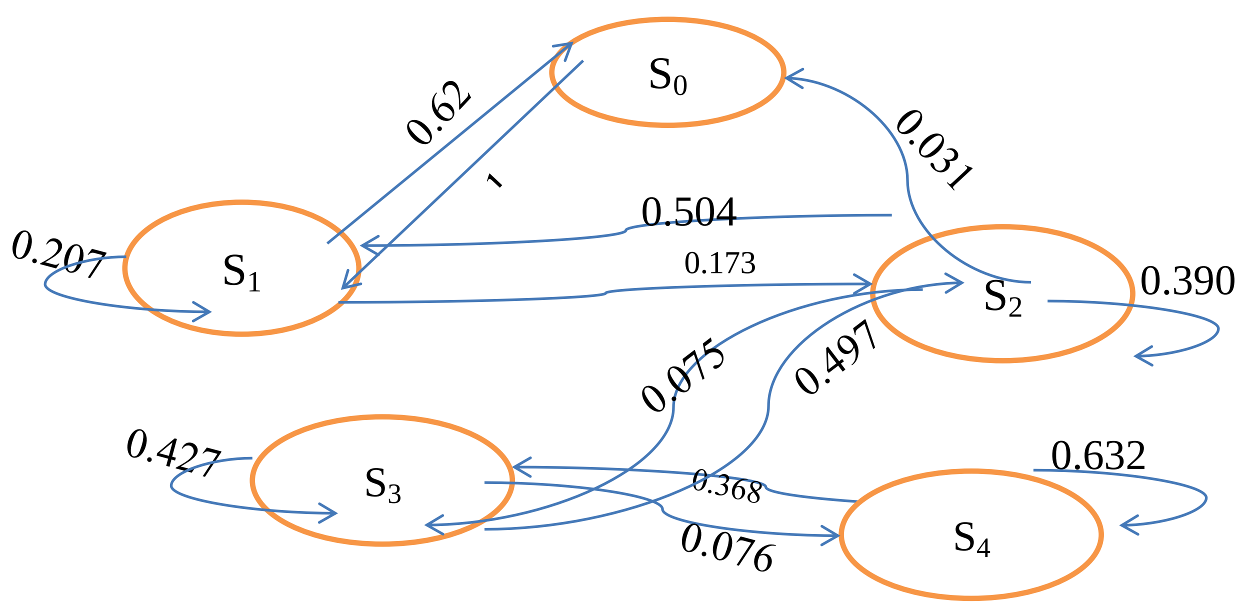
S_i / S_{i-1}	S_0	S_1	S_2	S_3	S_4	مجموع الصفوف
S_0	0	100	0	0	0	100
S_1	341	114	95	0	0	550
S_2	12	195	151	29	0	387
S_3	0	0	72	62	11	145
S_4	0	0	0	39	67	106
مجموع الأعمدة	353	409	318	130	78	

ومن الجدول (5) نقسم عناصر كل صف على المجموع الكلي للصفوف الذي تقع فيه ، فينتج عن ذلك مصفوفة عشوائية تستخدم كمصفوفة التحولات الاحتمالية للسنوات (2012–2009) ، وهي تعكس لنا الافتراضات الأولية عن حركة المرضى إذ إن S_4 ، S_0 حالتان منتهيتان ، وحصلنا على الجدول الآتي (الجدول 6) .

الجدول (6) يبين مصفوفة الانتقال لأعداد المصابين بذات الرئة

	S_0	S_1	S_2	S_3	S_4
S_0	0	1	0	0	0
S_1	0.62	0.207	0.173	0	0
S_2	0.031	0.504	0.390	0.075	0
S_3	0	0	0.497	0.427	0.076
S_4	0	0	0	0.368	0.632

ويمكن تمثيل مصفوفة الانتقال بمخطط شجيري (Digraph) يوضح صفات متسلسلة ماركوف المتمثلة بالمصفوفة P .



الشكل (1) يوضح المخطط الشجيري لمصفوفة الانتقال .

1-4: الثبوتية (Ergodic):

سوف نحاول في هذه الفقرة أَن نبين هل إذا كانت متسلسلة ماركوف المتمثلة بأعداد المصابين لمرض ذات الرئة ثبوتية أم لا . وهذا يتطلب إثبات مصفوفة الانتقال بدائية غير قابلة للتجزئة .

ومن ملاحظة المبرهنة (1)، والشكل (1) نجد أن مصفوفة الانتقال غير قابلة للتجزئة ، لأنها متصلة وممكن الوصول من أية حالة إلى جميع الحالات ، تحتوي مجموعة مغلقة واحدة. ولكي نثبت أن المصفوفة بدائية علينا إيجاد القيم الذاتية للمصفوفة P .واثبات أن إحدى هذه القيم تساوي واحداً وبقية القيم أَقل منها بالقيمة المطلقة . ولإيجاد القيم الذاتية نتبع حل النظام

$$|P - \lambda I| = 0$$

وبحل النظام في أعلاه حصلنا على القيم الذاتية الآتية:

$$\lambda_1 = -0.7179$$
$$\lambda_2 = 0.1699$$
$$\lambda_3 = 1$$
$$\lambda_4 = 0.4717$$
$$\lambda_5 = 0.7413$$

ومن هذه القيم نجد أن

$$\lambda_3 > |\lambda_i|, \qquad i = 1,2,4,5$$

نستنتج من ذلك أن مصفوفة الانتقال بدائية .

4−2: التوزيع المستقر (Stationary Distribution):

بما أن مصفوفة الانتقال تم إثباتها أنها ثبوتية (Ergodic) ، فإنه يوجد توزيع مستقر ووحيد المبرهنة (1)

ويمكن إيجاد التوزيع المستقر من حل النظام الآتي :

$$\pi p = \pi$$
$$\sum_{i=0}^{4} \pi i = 1$$

إذ إن $\pi_i = pr(state(s_i))$

ومن حل النظام في أعلاه حصلنا على القيم الآتية :

$$\pi_o = 0.3160$$
$$\pi_1 = 0.5020$$
$$\pi_2 = 0.1539$$
$$\pi_3 = 0.0232$$
$$\pi_4 = 0.0048$$

من هذه القيم نستنتج أن احتمالية الإصابة بهذا المرض تكون أكبر في أعمار الأشخاص دون سن الرابعة عشرة وتضعف احتمالية الإصابة كلما ازداد العمر وهذه نتيجة منطقية. وإن احتمالية انتقال

المريض الى حالة الشفاء أيضاً مقبولة المتمثلة بقيمة $\pi_o = 0.3160$

الاستنتاجات :

من خلال دراستنا للسلسلة الزمنية لعدد الإصابات بمرض ذات الرئة في محافظة نينوى تبين أن المتسلسلة ثبوتية . وأن احتمالية الإصابة بهذا المرض تكون أكبر في أعمار الأشخاص أَقل أو يساوي (14) سنة ، وتضعف الاحتمالية بالإصابة إذا ازداد عمر الشخص عن (14) سنة وهذه نتائج معقولة مما يشير إلى كفاءة الطريقة والنموذج المستخدم في هذه الدراسة .

المصادر:

1- ذنون ، باسل يونس (2011) " النمذجةالماركوفية مع تطبيقات عملية الجزء الأول الأساسيات " رقم الإيداع في دار الكتب والوثائق ببغداد 1145 .
2- ذنون ، باسل يونس (1991) " النمذجة الاحتمالية والمتغيرات العشوائية " دار ابن الأثير للطباعة والنشر .
3- العبيدي ، شهاب احمد إبراهيم (2007) " سلسلة ماركوف الثبوتية مع التطبيق " رسالة ماجستير ، جامعة تكريت .
4- الصميدعي ، وفاء محي الدين (2000) " تحليل ونمذجة السلسلة الزمنية لدرجات الحرارة في مدينة الموصل " رسالة ماجستير ، جامعة تكريت .
5- العذارى ، فارس مسلم والوكيل ، علي عبد الحسين (1991) " العمليات التصادفية " الطبقة الأولى ، وزارة التعليم العالي والبحث العلمي ، جامعة بغداد .
6- Dean, L. Isaacson (1976) “ Markov chain theory and application” John Wiley and Sons, Inc .

7- Salim A.J and thanoon , B.Y. (1984) “ A Markov chain model for river flow “ Qatar unin , Sci, j, 16(2), PP (231-235).

8- Travers L. V and Souza, R.C. (1988) “ A multiple distribution markovian model for annual hydrologic tame series , stochastic hydrology, 2, PP (292-302).