مختارات أساسية

مفاهيم أساسية:

ان دراسة المحاكاة وأساليبها تعتمد اعتماداً كبيراً على مفاهيم أساسية في كل من نظرية الاحتمال Probability Theory وعلم الإحصاء Statistics. ولأجل ان يكون هذا الكتاب متكاملاً ذاتياً وبحيث يجد القارئ الأساسيات التي يحتاجها من اجل فهم الأفكار المطروحة، فقد ارتأينا وضع هذا الفصل كي يستعرض وبإيجاز المفاهيم الأساسية في نظرية الاحتمال وعلم الإحصاء والتي تعتمد عليها الفصول القادمة.

ان نظرية الاحتمال تتعامل مع **اللايقين Uncertainty** في حدوث الأحداث Events المختلفة من حيث توقع حدوثها. واللايقين يعتمد أساسا على نقص في المعلومات المتاحة مما يؤدي بالنتيجة إلى الحاجة إلى قياس احتمالية حدوث الأحداث ذات العلاقة، واللايقين يتناسب طرديا مع كمية المعلومات المحجوبة. فعندما يقول الطالب "احتمال ان انجح في الامتحان ٦٠%"، فيجب الانتباه جيدا إلى ان عدم يقين الطالب بنتيجة الامتحان ناجم عن جهله بجميع الظروف التي تحدد درجته في الامتحان (طريقة إجابته وظروف تصحيح الإجابات...الخ)، علما ان نتيجة الطالب الفعلية لا وجود للايقين فيها، فهي نتيجة حتمية موجودة في اللجنة الامتحانية. كذلك نستنتج من هذا المثال ان اللايقين حول مسألة ما يكون لفئة أو طائفة معينة وليس بالضرورة للجميع. فنتيجة الامتحان مجهولة للطلبة، في حين أنها معلومة للمدرس أو لأعضاء اللجنة الامتحانية.

ان الأحداث بعامة، تقسم على ثلاثة أنواع رئيسة:

١- أحداث مؤكدة الحدوث Certain Events: وهي أحداث يُجزم بحدوثها يقينا. وسبب جزمنا هو ان هذه الأحداث قد حدثت بالتأكيد في جميع الفترات الزمنية الماضية، مثل حدث "موت الإنسان" هو حدث مؤكد لأننا لم نلحظ أي إنسان قد استطاع الخلود منذ الأزل.

٢- أحداث مستحيلة الحدوث Impossible Events: وهي أحداث نجزم بعدم حدوثها ولا نتوقع ان تحدث على الإطلاق، مثل حدث " ان يعيش الإنسان بدون ماء أو هواء". وتجدر الإشارة إلى أننا حين نقول ان حدثاً ما هو حدث مؤكد أو مستحيل فإننا نستدل على ذلك من خلال المعلومات التي يجهزنا بها حاضر ذلك الحدث وماضيه. أما إذا حدثت أمور غيبية أو معجزات فتلك أمور خارج نطاق دراستنا لان المعجزة هي عبارة عن خرق للقوانين المعروفة.

محاضرات حاسوبية مرحلة ثانية-مدرس المادة: د.رشدي عبدالله

٣- أحداث محتملة (غير مؤكدة) Probable (Uncertain Events): وهي أحداث لا يمكن ان نجزم جزماً مؤكداً بحدوثها أو عدم حدوثها. أي أنها أحداث غير مؤكدة الحدوث ولا مستحيلة الحدوث. من أمثلة هذه الأحداث "سقوط الأمطار يوم غد" ، "نجاح الطالب في الامتحان" ، "عطب جهاز معين".

ان **التجربة Experiment** يمكن تعريفها بأنها حالة يمكننا من خلالها ملاحظة نتيجة ما. والتجربة إما ان تكون **حتمية Deterministic**، بحيث إذا أعيدت أكثر من مرة واحدة وفي الظروف نفسها في كل مرة فإنها تعطي دائما وفي الظروف نفسها النتيجة نفسها. أو أن تكون **التجربة عشوائية Random** ، بحيث إذا أعيدت أكثر من مرة واحدة وفي الظروف نفسها فإنها في كل مرة ممكن ان تعطي نتائج مختلفة، ولو بمقدار طفيف عن المرات السابقة.

ان وصف عائد (حاصل/ناتج) التجربة العشوائية يتم عادة من خلال متغيرات يطلق عليها **متغيرات عشوائية Random Variables. والمتغير العشوائي** هو عبارة عن دالة رياضية (تطبيق Mapping) ينظم عائد التجربة العشوائية إلى أعداد، ويرمز عادة للمتغيرات العشوائية بالحروف اللاتينية الكبيرة مثل X و Y و Z. والمتغيرات العشوائية تكون بأحد النوعيين الآتيين:

١-المتغيرات العشوائية المتقطعة Discrete Random Variables: وهي المتغيرات التي تأخذ قيماً منفصلة، عادة يعبر عنها بشكل أعداد صحيحة، ٠، ١، ٢،....

٢-المتغيرات العشوائية المستمرة Continuous Random Variables: وهي المتغيرات التي تأخذ قيماً متصلة، يعبر عنها بشكل أعداد حقيقية.

ان السلوك (القانون) الرياضي لاحتمالات جميع القيم الممكنة للمتغير العشوائي يسمى **بالتوزيع الاحتمالي Probability Distribution**. ان **التوزيع الاحتمالي** هو عبارة عن كيان احتمالي معين له مزاياه وخصائصه الخاصة به. وهناك العديد من التوزيعات الاحتمالية الأساسية المعروفة، كما ان كل ظاهرة يكون سلوكها الاحتمالي وفق توزيع احتمالي معين. لذا نجد ان كل توزيع احتمالي يكون مناسبا لمجموعة كبيرة من المتغيرات المختلفة في أصولها ومنشئيها إلا أنها متشابهة في سلوكها الاحتمالي.

المعدل (الوسط الحسابي) Mean للمتغير X ويرمز له عادة μ.

التباين Variance المتغير X ويرمز له عادة Var(X) أو σ^2.

محاكاة حاسوبية_ مرحلة ثانية_ مدرس المادة: د. رشدي عبدالله

التغاير والارتباط Covariance and Correlation:

عندما يراد دراسة العلاقة بين متغيرين عشوائيين، تظهر الحاجة لدراسة عزم معين يعرف **بالتغاير Covariance**. فإذا كان المتغير X متغيرا عشوائيا معدله μ_x وانحرافه المعياري σ_x وكان Y متغيراً عشوائيا آخر معدله μ_y وانحرافه المعياري σ_y، فان تغاير المتغيرين X و Y يرمز له $Cov(X,Y)$ ويعرف بالشكل الآتي:

$$Cov(X,Y)=E[(X-\mu_x)(Y-\mu_Y)].$$

ان وحدة التغاير هي حاصل ضرب وحدة المتغير X بوحدة المتغير Y، وللتخلص من الوحدات يُعمد إلى تحويل التغاير إلى مقياس مجرد من الوحدات، وذلك بقسمة التغاير على كل من الانحراف المعياري للمتغير X والمتغير Y كي ينتج المقياس المعروف **بمعامل الارتباط Correlation Coefficient** والذي يرمز له ρ ، أي ان:

$$\rho=\frac{\text{cov}(X,Y)}{\sigma_X\sigma_Y}\quad .$$

ومن الخصائص المهمة لمعامل الارتباط انه مقياس مجرد من الوحدات، وان قيمته العددية محصورة بين (١−) و (١+) . ان معامل الارتباط يفيد في قياس كل من قوة العلاقة واتجاهها بين المتغيرين X و Y. ان الإشارة الموجبة لمعامل الارتباط تعني ان العلاقة بينهما طردية (زيادة في احدهما يصاحبها زيادة في قيمة الآخر)، أما الإشارة السالبة لمعامل الارتباط فتعني بان العلاقة بين المتغيرين هي علاقة عكسية (زيادة احدهما يصاحبها نقصان بالآخر والعكس بالعكس). وأما قوة العلاقة فيمكن قياسها من خلال القيمة المطلقة لمعامل الارتباط، فكلما اقتربت القيمة المطلقة لمعامل الارتباط من الواحد ازدادت قوة العلاقة.

ومن المفاهيم التي لها أهمية كبيرة في المجالات التطبيقية المختلفة ما يعرف **بالمتسلسلات الزمنية**(*) **Time Series** (أو **دالة زمن Time Function**، أو **الإشارة Signal**). وبإيجاز شديد فان أي ظاهرة تتغير بتغير دليل آخر (كالزمن مثلا) تسمى متسلسلة زمنية. وكأمثلة على المتسلسلات الزمنية الأسعار اليومية للسلع والبضائع، ودرجات الحرارة اليومية في منطقة معينة، والإنتاج الشهري لمصنع معين، والأرباح السنوية لشركة معينة.

محاكاة حاسوبية_مرحلة ثالثة مدرس المادة: د.شذى عبدالله

(*)هناك مصطلحان علميان يُعرّبان في المصادر العربية المتداولة "**سلسلة**": الاول هو Series والثاني هو Chain. وحسب معجم الرياضيات لسنة ١٩٩٥ لمؤلفيه بوروفسكي و بورفان، ترجمة علي مصطفى بن الأشهر، فان مصطلح Chain يُعرّب إلى "**سلسلة**"، في حين ان مصطلح Series يُعرّب إلى "**مُتسلسلة**". وبذلك يكون تعريب Markov Chains هو "**سلاسل ماركوف**"، في حين ان تعريب Time Series هو "**متسلسلة زمنية**" للمفرد و "**متسلسلات زمنية**" للجمع. علما بان التعريب المتداول للمصطلح Time Series هو "**سلاسل زمنية**".

ومن الخصائص البنيوية للسلاسل الزمنية وجود ترابط داخلي بين قيم السلسلة الزمنية. فقيمة السلسلة الزمنية في فترة زمنية معينة لها علاقة مع قيمها في الفترات الزمنية السابقة. إن التغاير بين قيمة المتسلسلة الزمنية $\{X_t\}$ عند الزمن t، أي X_t، وقيمتها عند الزمن t-k، أي X_{t-k} يرمز له $\gamma(k)$، اذ ان العدد الصحيح k يسمى عادة **بفترة الإبطاء Lag**. وباعتبار $\gamma(k)$ دالة بدلالة فترة الإبطاء k فان $\gamma(k)$ تسمى عادة **بدالة التغاير الذاتي Autocovariance Function** للمتسلسلة الزمنية $\{X_t\}$.

خصائص دالة التغاير الذاتي:

١- التغاير الذاتي عند فترة الإبطاء ٠ هو بعينه التباين، أي ان $\gamma(٠)=Var(X_t)$.

٢- دالة التغاير الذاتي متناظرة حول الصفر (للمتسلسلات الزمنية ذوات القيم الحقيقية). أي ان $\gamma(+k)=\gamma(-k)$ لجميع قيم k.

٣- لجميع قيم k فان $|\gamma(k)| \le \gamma(0)$.

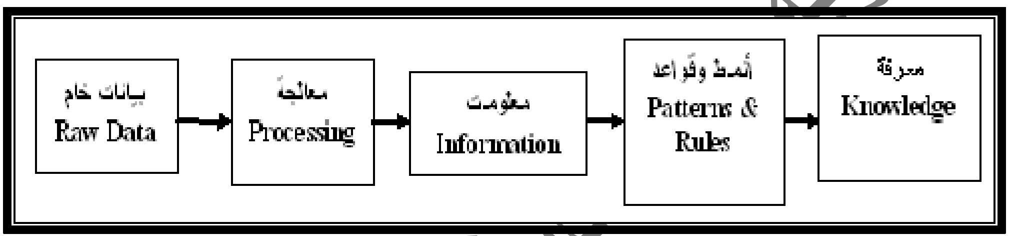
وبالاستفادة من الخاصة الأخيرة تُحوّل دالة التغاير الذاتي إلى دالة معيارية تُعرف **بدالة الارتباط الذاتي Autocorrelation Function** والتي تمثل معامل الارتباط بين X_t و X_{t-k}، و يرمز إلى دالة الارتباط الذاتي $\rho(k)$ وتعرّف على النحو الآتي:

$$\rho(k)=\frac{\gamma(k)}{\gamma(o)} \qquad ; k=١,٢,\ldots$$

محاضرة حاسوبية - مرحلة ثانية - مدرس المادة: د.شذى عبدالله

٤-٣ مختارات من علم الإحصاء:

ان **البيانات Data** (أو البيانات الخام Raw Data) هي حقائق أولية أو تعاملات أو أشكال مجمعة بطريقة القياس أو المشاهدة أو من خلال سجلات محفوظة. ان هذه البيانات تتضمن معلومات قيّمة، إلا ان هذه المعلومات لا تكون ظاهرة للعيان. ان الحقائق Facts والميزات features أشبه ما تكون بالأحجار الكريمة الخام الموجودة في ركام الأتربة والصخور، وهذه الأحجار الكريمة لا تبرز قيمتها إلا بعد تنقيتها ومعالجتها. ومن خلال معالجة البيانات الخام Raw Data باستخدام الأساليب الإحصائية تُحوّل البيانات إلى معلومات Information ، ثم من خلال أنماط وقواعد معينة تُحوّل المعلومات إلى معرفة Knowledge وكما هو مبين في المخطط الآتي:



الشكل (٤-١): البيانات والمعلومات والمعرفة.

يمكن تقسيم البيانات بعامة إلى نوعين من حيث البنية التركيبية لها:

أ- **بيانات إحصائية Statistical Data**: وهي ما يطلق عليها عادة تسمية **عينة عشوائية Random Sample**. تكون البيانات الإحصائية ذات استقلالية ذاتية وتتميز بعدم وجود ترابط داخلي فيما بينها، مثل أوزان مجموعة من الأشخاص أو أعمار مجموعة من الأجهزة.

ب- **بيانات المتسلسلات الزمنية Time Series Data**: المتسلسلة الزمنية هي متتالية (متتابعة) بيانات مُفهرسة (مُدلّلة) زمنيا، وتتضمن غالبا قراءات (قياسات) عددية متباعدة زمنيا بانتظام، مثلا كل دقيقة أو كل يوم أو كل شهر أو كل سنة. ان القيم العددية (البيانات) للمتسلسلات الزمنية سوف نطلق عليها تسمية **المتسلسلة المُتَحَقِقَة** (**) **Realization**، أو

(**) ان مجموعة جميع السجلات Records الممكنة من متسلسلة زمنية تسمى **Ensemble**، وسوف نُعَرّب هذا المصطلح بالمصطلح **مجموعة السجلات**، وان أي سجل مُدرك مُتحقق من تلك السجلات يسمى بالمصادر الأجنبية **Realization**. ونظرا لعدم ورود مصطلح مُعرّب لهذا المصطلح الأجنبي فإننا نقترح تعريبه **بالسجل المُتحَقِقْ**، والمتسلسلة الزمنية التي تَتَحَقق وتظهر كحقيقة واقعة **بالمتسلسلة المُتَحَقِقَة**، (المؤلف).

اختصارا **المُتَحَقِقَة**. إن ما يميز المتسلسلة المُتحققة عن البيانات الإحصائية هو عدم وجود استقلالية ذاتية بين عناصرها بسبب وجود ترابط ذاتي فيما بينها. وكأمثلة على المتسلسلات الزمنية، درجات الحرارة اليومية لمنطقة معينة، والمبيعات الشهرية لسلعة معينة ، والأرباح السنوية لشركة معينة.

محاكاة حاسوبية – مرحلة ثانية – مدرس المادة: د.رشى عبدالله

٤–٣–١ التقدير Estimation:

لنفرض ان X و Y متغيران عشوائيان معدلاهما μ_X و μ_Y على التوالي، وتبيانهما σ_X^2 و σ_Y^2 على التوالي، وان معامل الارتباط بينهما هو ρ . ولو كانت $\underline{x}=\{x_1, x_2,..., x_n\}$ و $\underline{y}=\{y_1, y_2,..., y_n\}$ هما عينتان عشوائيتان من المتغيرين X و Y على التوالي. فان المعدلين μ_X و μ_Y يمكن تقديرهما من البيانات على النحو الآتي:

$$\hat{\mu}_x = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \hat{\mu}_y = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

أما التباينين σ_X^2 و σ_Y^2 فيمكن تقديرهما على النحو الآتي:

$$\hat{\sigma}_x^2 = S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 .$$

$$\hat{\sigma}_y^2 = S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}^2 .$$

وتجدر الملاحظة انه لأسباب إحصائية معينة يفضل قسمة مقدّري التباين على (n-١) بدلا من (n)، إلا أننا سوف نستخدم المقدرين السابقين، مع مراعاة أننا في هذا الكتاب سوف نتعامل مع عينات كبيرة.

أما تغاير المتغيرين X و Y فيمكن تقديره باستخدام المقدّر الآتي:

$$\hat{\gamma} = \mathrm{cov} = \frac{1}{n}\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y} .$$

وبذلك يكون مقدّر معامل الارتباط بين المتغيرين هو:

$$\hat{\rho} = r = \frac{\hat{\gamma}}{\hat{\sigma}_x \hat{\sigma}_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2][\sum_{i=1}^{n} y_i^2 - n\bar{y}^2]}}$$

لنفرض الآن ان $\underline{x} = \{x_1, x_2, ..., x_n\}$ هي متسلسلة مُتحققة مأخوذة من متسلسلة زمنية معينة $\{X_t\}$. ان المعدل والتباين لهذه المتسلسلة يمكن تقديرهما تماما كما في حالة العينة العشوائية وذلك باستخدام المقدرين $\hat{\mu}_X$ و $\hat{\sigma}_X^2$ على التوالي. أما دالة التغاير الذاتي فيمكن تقديرها باستخدام المقدّر الآتي:

$$\hat{\gamma}(k) = c(k) = \frac{1}{(n-k)} \sum_{i=k+1}^{n} (x_i - \bar{x})(x_{i-k} - \bar{x}) \qquad ; k=٠,١,\dots,(n-١)$$

وأما دالة الارتباط الذاتي فان مقدرها هو

$$\hat{\rho}(k) = r(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \qquad ; k=٠,١,\dots,(n-١)$$

وتجدر الملاحظة انه من الناحية العلمية يفضل ان لا تزيد قيمة k عن $\frac{n}{2}$ عند تقدير دالة التغاير الذاتي أو دالة الانحدار الذاتي.

مثال:

افرض ان

$\underline{x} = \{٥,٧,٤,٢,٩,٧,١٠,٩,٩,٨\}$

$\underline{y} = \{١٢,١٦,١١,١٨,٢٢,١٩,١٨,٢٢,١٧,١٥\}$

اوجد المعدل والتباين والانحراف المعياري لكل من $\underline{x}$ و $\underline{y}$ وكذلك التغاير ومعامل الارتباط بينهما.

الحل:

النتائج:

MX =
٧

MY =
١٧

محاضرات حاسوبية _ مرحلة ثانية-مدرس المادة: م.د. بشرى عبدالله

SX =
٢.٥٨٢٠

SY =
٣.٦٨١٨

co =

٦.٦٦٦٧ ٤.٨٨٨٩
٤.٨٨٨٩ ١٣.٥٥٥٦

r =
١.٠٠٠٠ ٠.٥١٤٣
٠.٥١٤٣ ١.٠٠٠٠

من المصفوفتين الأخيرتين نستنتج بان: Var(X)=٦.٦٦٦٧ و Var(Y)=١٣.٥٥٥٦ و Cov(X,Y)=٤.٨٨٨٩ ومعامل الارتباط r=٠.٥١٤٣.

المثال

اوجد المعدل والانحراف المعياري ودالة الارتباط الذاتي للمتسلسلة المُتحققة الآتية:

$\underline{x}$ ={٧,٩,١٤,١٢,٨,٦,٨,٨,١٠,١٤,١٢,١١,٨,٩,١٥}.

الحل:

النتائج:

MX =
١٠.٠٦٦٧

SDX =
٢.٧٨٩٤

Bounds =
٠.٥١٦٤ -٠.٥١٦٤

k	r(k)
٠	١.٠٠٠٠

محاكاة حاسوبية _ مرحلة ثانية - مدرس المادة: د. بشرى عبدالله

١.٠٠٠٠	٠.٢٥٧٦
٢.٠٠٠٠	-٠.٣٤٠٤
٣.٠٠٠٠	-٠.٣٣٩٢
٤.٠٠٠٠	-٠.١٦٣٠
٥.٠٠٠٠	٠.١١٢٤
٦.٠٠٠٠	٠.١٣٠٧
٧.٠٠٠٠	٠.٢٢١٣
٨.٠٠٠٠	٠.٠٣٥٢
٩.٠٠٠٠	-٠.٢٩٦٦
١٠.٠٠٠٠	-٠.٢٥٠٧
١١.٠٠٠٠	٠.٠٤٣٠
١٢.٠٠٠٠	٠.٢٤٦٨
١٣.٠٠٠٠	-٠.٠١٨٣
١٤.٠٠٠٠	-٠.١٣٨٩

محاكاة حاسوبيه_مرحلة ثانية-مدرس المادة: د.شذى عبدالله