الفصل الثالث

بعض نماذج الاحتمال

SOME PROBABILITY MODELS

كما أشرنا فى نهاية البند (١ – ٣) ، نصف المتغير سـ بأنه **متغير عشوائي** إذا كانت قيمه أعداداً حقيقية يتأثر قياسها بعوامل عشوائية ويكون ظهورها مصحوباً باحتمالات محددة . ولكل متغير عشوائي توزيع يربط بين قيمه واحتمالاتها يسمى بتوزيع الاحتمال لهذا المتغير . وتنقسم توزيعات الاحتمال للمتغيرات العشوائية إلى نوعين هما توزيعات الاحتمال الوثابة وتوزيعات الاحتمال المتصلة بحسب كون المتغير من النوع الوثاب أو من النوع المتصل .

(٣ – ١) توزيعات الاحتمال الوثابة

DISCRETE PROBABILITY DISTRIBUTIONS

ليكن سـ متغيراً حقيقياً وثاباً يأخذ القيم

$س_{١}$ ، $س_{٢}$ ، $س_{٣}$ ، ... باحتمالات قدرها

$ل_{١}$ ، $ل_{٢}$ ، $ل_{٣}$ ، ... حيث مجـ $ل_{ر}$ = ١

إن تجمع قيم سـ مع احتمالاتها المناظرة في مجموعة من الأزواج المرتبة

{ ($س_{١}$ ، $ل_{١}$) ، ($س_{٢}$ ، $ل_{٢}$) ، ($س_{٣}$ ، $ل_{٣}$) ، ... } .

يسمى بتوزيع الاحتمال للمتغير العشوائي سـ .

وفي كثير من الأحيان يمكن أن نجد داله غير سالبة د حيث :

د (س) = ل (سـ = س) (١)

أى حيث تكون قيمة الدالة د عند س مساوية لاحتمال أن يأخذ المتغير سـ القيمة س . وتسمى هذه الدالة حينئذ **بدالة كتلة الاحتمال للمتغير** سـ probability mass function

مثال (٣ - ١) :

التوزيع الآتي هو توزيع الاحتمال لعدد مرات إصابة الهدف لجندى ما يطلق بندقيته على هدف ثابت ٥ مرات .

عدد مرات إصابة الهدف س :	صفر	١	٢	٣	٤	٥
احتمال هذا العدد ل :	صفر	٠,٤	٠,٣	٠,٢	٠,١	صفر

لاحظ أن مجـ ل = ٠ + ٠,٤ + ٠,٣ + ٠,٢ + ٠,١ + ٠ = ١

ويمكن هنا التعبير عن الاحتمالات بواسطة الدالة د المعرفة كالآتي :

$$د (س) = \frac{٥ - س}{١٠} \quad حيث س = ١ ، ٢ ، ٣ ، ٤ ، ٥$$

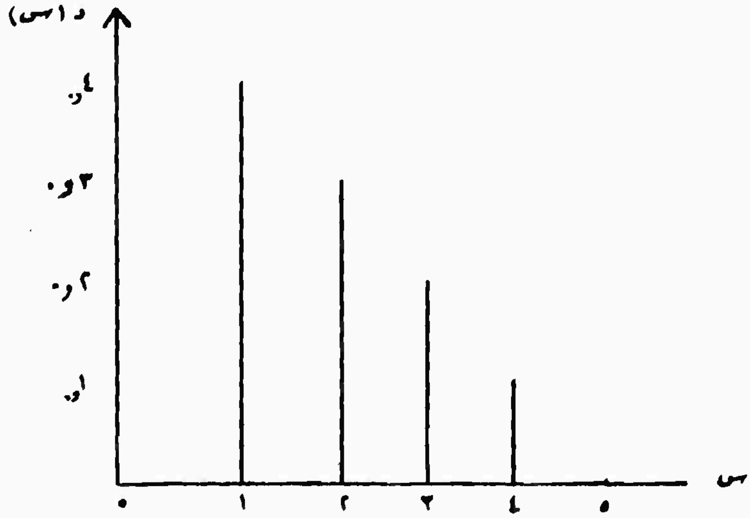
$$، صفر \quad فيما عدا ذلك$$

إذ أنه بوضع س = ١ ، ٢ ، ٣ ، ٤ ، ٥ نحصل على الاحتمالات ٠,٤ ، ٠,٣ ، ٠,٢ ، ٠,١ ، صفر .

ونمثل هذه الدالة أو هذا التوزيع بيانيا كما في الشكل (٣ - ١) الآتي .

(٣-١-١) الوسط الحسابي والتباين : MEAN AND VARIANCE

إن توزيعات الاحتمال الوثابة تشبه التوزيعات التكرارية ، إلا أن الاحتمالات ل تحل محل التكرارات ك ، كما أن حجم التوزيع هو الواحد الصحيح دائماً . وهذا الواحد يشير إلى أن هناك احتمالا قدره الواحد الصحيح موزعاً على القيم المختلفة للمتغير ولذلك سمى التوزيع بتوزيع الاحتمال .



الشكل (٣ - ١) الأعمدة البيانية لتوزيع احتمال
عدد مرات إصابة الهدف لجندى يطلق على الهدف خمس طلقات

ومن هنا كان تعريفا الوسط الحسابي والتباين لتوزيع احتمال - وسنرمز لهما بالرمزين μ ، σ^2 ، يشبهان تعريفي الوسط الحسابي والتباين للتوزيع التكرارى فهما يعرفان كالآتي :

الوسط الحسابي = $\mu = \text{مجـ}\ \text{ل}_{\text{ر}}\ \text{س}_{\text{ر}}$ (٢)

التباين = $\sigma^2 = \text{مجـ}\ \text{ل}_{\text{ر}}\ (\text{س}_{\text{ر}} - \mu)^2$

أو = $\text{مجـ}\ \text{ل}_{\text{ر}}\ \text{س}^2_{\text{ر}} - \mu^2$ (٣)

أما الانحراف المعيارى فهو بالطبع الجذر التربيعى للتباين .

ففي المثال (٣ - ١) السابق نجد أن :

$$\mu = \text{مجـ}\ \text{ل}_{\text{ر}}\ \text{س}_{\text{ر}}$$

$$= 0\times0 + 0.4\times1 + 0.3\times2 + 0.2\times3 + 0.1\times4 + 0\times5 = 2$$

، $\sigma^2 = \text{مجـ}\ \text{ل}_{\text{ر}}\ \text{س}^2_{\text{ر}} - \mu^2$

$$= (0\times0 + 0.4\times1 + 0.3\times4 + 0.2\times9 + 0.1\times16 + 0\times25) - 4$$

$$= 5 - 4 = 1$$

(٣ -٢) نماذج الاحتمال : PROBABILITY MODELS

نموذج الاحتمال لمتغير عشوائي س هو توزيع احتمال ذو صيغة رياضية محددة يفترض أنها تعكس سلوك المتغير س . ويعبر عن الاحتمالات من هذا النموذج بدلالة واحد أو أكثر من أدلة مجهولة تتوقف على خواص المجتمع وطريقة المعاينة منه . ويبني كل نموذج احتمال على افتراضات خاصة تصور تصويراً مناسباً الميكانيكية العشوائية التي تسبب الاختلافات في مشاهداتنا عن المتغير .

وسنتناول فيما يلي أربعة من أشهر نماذج الاحتمال للمتغيرات العشوائية الوثابة تعرف بتوزيع ذى الحدين وتوزيع بواسون وتوزيع باسكال والتوزيع الهندسي .

(٣ - ٣) توزيع ذى الحدين : THE BINOMIAL DISTRIBUTION

في كثير من الأحيان يكون اهتمامنا منصباً على وجود أو عدم وجود خاصة ما في وحدات أو عناصر مجتمع ما . ولذلك ننظر إلى المجتمع على أنه مقسم إلى قسمين منفصلين بحسب هذه الخاصة . فمثلا قد نقسم مجتمعاً من الطلاب بحسب خاصة القومية إلى القسمين : عربي / غير عربي أو بحسب الجنس إلى : ذكر/ أنثي أو إلى القسمين : يدخن / لا يدخن أو إلى القسمين : يلبس نظارة / لا يلبس نظارة أو إلى أى قسمين منفصلين متكاملين .

في مثل هذه الحال تتركز الصفة الأساسية للمجتمع في دليل أو بارامتر واحد هو نسبة أى من القسمين - وليكن القسم الأول - إلى المجتمع كله ، وسنرمز إلى هذه النسبة بالرمز ح . فمثلا قد تكون ح نسبة الطلاب العرب في المجتمع ، وهنا تكون نسبة الطلاب غير العرب هى ١ - ح = ك مثلا . وهذا يعني أننا إذا سحبنا عشوائياً عنصراً من المجتمع فإن ح تعبر عن احتمال أن يكون هذا العنصر من القسم الأول (طالب عربي مثلا) كما أن ك تعبر عن احتمال أن يكون العنصر من القسم الثاني (طالب غير عربي) .

سنسمى ظهور عنصر من القسم الأول **نجاحاً** للخاصة أو الحدث الذى ندرسه وظهور عنصر من القسم الثاني **فشلا** للحدث ، أى سنعتبر أن لدينا حدثاً واحداً – مثلا ظهور طالب عربي – إما أن يقع أو لا يقع .

إن هدفنا الأساسي من هذه الدراسة يتلخص فيما يلي : نفرض أننا سحبنا من المجتمع عينة عشوائية حجمها ن = ١٠ مثلا . هناك ١١ حالة ، إذ يمكن أن تكون هذه العينة خالية من أى عنصر من القسم الأول كما يمكن أن تشتمل على عنصر واحد فقط من هذا القسم أو تشتمل على عنصرين أو ثلاثة أو أربعة أو ... عشرة . أى أن عدد مرات نجاح الحدث يمكن أن يكون ٠ ، ١ ، ٢ ، ٣ ، ... ، ١٠ . والسؤال الذى نستهدف الإجابة عنه هو : ما احتمال كل من هذه الحالات ؟ وبمعني آخر إذا اعتبرنا أن لدينا متغيراً عشوائياً س يعبر عن **عدد مرات نجاح الحدث** فما توزيع الاحتمال لهذا المتغير ؟

إن الإجابة عن هذا السؤال تتوقف على ما نضعه من افتراضات نرى أنها مناسبة لما يهمنا من أوضاع ويمكن تحققها في عملية التجريب . وسوف نتبني هنا الافتراضات أو الشروط الآتية :

(١) عشوائية العينة :

سنفترض أن المعاينة (أى سحب العناصر من المجتمع) عشوائية .

(٢) ثبات الدليل ح :

سنفترض أن احتمال نجاح الحدث هو عدد **ثابت** ح طوال عملية سحب العينة . ولكى يتحقق هذا الفرض سنعتبر أن حجم المجتمع كبيراً بالنسبة لحجم العينة وبالتالى فإنه حتي إذا كانت المعاينة بغير إرجاع يكون التغير الذى يحدث في قيمة ح تغيراً طفيفاً يمكن التجاوز عنه ، كما سنفترض أيضاً أن قيمة ح لا تختلف من عينة إلى أخرى .

(٣) استقلال الأحداث :

سنفترض أن نجاح (أو فشل) الحدث في أى سحبة مستقل عما نتج من نجاح أو فشل في السحبات السابقة ، أى أن ما تسفر عنه أى سحبة لا يتأثر بأى حال بما نتج في السحبات الأخرى . كما سنفترض أن عدد مرات نجاح الحدث في عينة ما مستقل عن عدد مرات نجاحه في أى عينة أخرى .

تحت هذه الشروط وباستخدام قاعدة الضرب للأحداث المستقلة وقاعدة الجمع للأحداث المتنافية – انظر البند (١ – ٧) عن توافقات الاحتمال – نستطيع أن نثبت رياضياً أنه إذا كان المتغير س يعبر عن عدد مرات نجاح الحدث، فإن دالة كتلة احتماله تأخذ الصورة الآتية :

$$\text{د}(\text{س}) = \text{ق}^{\text{ن}}_{\text{س}}\ \text{ح}^{\text{س}}\ \text{ك}^{\text{ن}-\text{س}} \quad \text{حيث س} = ٠ ، ١ ، \dots ، \text{ن} \qquad (٤)$$

وحيث ١ > ح > . ، ك = ١ – ح ، ن هو حجم العينة .

ويسمى توزيع الاحتمال حينئذ بتوزيع ذى الحدين .

كما نستطيع أن نثبت رياضياً أن

الوسط الحسابي للتوزيع = ن ح (٥)

تباين التوزيع = ن ح ك (٦)

وجدير بالملاحظة أن الدالة (٤) تتوقف على اثنين من الأدلة هما ن ، ح ومعرفة هذين الدليلين تحدد التوزيع تحديداً تاماً . ولذلك سنرمز لتوزيع ذى الحدين بالرمز حد (ن ، ح) .

ملاحظة :

يمكن إيجاد التوافقات $\text{ق}^{\text{ن}}_{\text{س}}$ بالطريقة الحسابية المعتادة أو باستخدام مثلث

باسكال أو باستخدام جداول جاهزة حسبت فيها هذه التوافقات – انظر الجدول (٢) في ملحق هذا الكتاب وهو يعطى التوافقات $\binom{ن}{س}$ لبعض قيم ن ، س (ن = ١ ، ٢ ، ٣ ، ... ، ٢٠ و س = ٠ ، ١ ، ٢ ، ... ، ٢٠) ، كما يمكن إيجاد قيم الاحتمالات (٤) من الجدول (٣) لبعض قيم ن ، ح (ن = ٢ ، ٣ ، ... ، ١٥ و ح = ٠,٠٥ ، ٠,١ ، ٠,٢ ، ٠,٣ ، ٠,٤ ، ٠,٥ ، ٠,٦ ، ٠,٧ ، ٠,٨ ، ٠,٩ ، ٠,٩٥) .

مثال (٣ – ٢) :

افرض أن احتمال ولادة مولود ذكر هو ٠,٥ واعتبر العائلات التي أنجبت ٤ أطفال .

(أ) أوجد توزيع احتمال المتغير س الذى يعبر عن عدد الذكور في العائلات ذوات الأربعة الأطفال .

(ب) إذا أخذنا عشوائياً ٢٠٠٠ عائلة من هذا النوع فما العدد الذى نتوقعه للعائلات التي يكون بها ولدين على الأقل ؟

الحـــل :

لدينا مجتمع من الولادات مقسم إلى القسمين ذكر / أنثي ، واحتمال وقوع أو نجاح الحدث « المولود ذكر » هو عدد **ثابت** ح = $\frac{1}{2}$. إن المتغير س يعبر هنا عن عدد الأولاد (الذكور) في العائلات ذوات الأربعة الأطفال أى عدد المرات التي تحدث فيها ولادة مولود ذكر في هذا النوع من العائلات وإذن فالمتغير س لا يأخذ إلا القيم ٠ ، ١ ، ٢ ، ٣ ، ٤ . إن كل عائلة تعتبر عينة **عشوائية** حجمها ن = ٤ . فإذا فرضنا أن إنجاب مولود ذكر (أو أنثي) في أى ولادة **مستقل** عما ينتج في الولادات الأخرى تكون الافتراضات الثلاثة لتوزيع ذى الحدين متوفرة ويكون للمتغير س توزيع ذى الحدين دليلاه ن = ٤ ، ح = $\frac{1}{2}$ وبالتالى تكون دالة كتلة احتماله :

$$د\,(س) = {}^{٤}ق_{س} \left(\frac{١}{٢}\right)^{س} \left(\frac{١}{٢}\right)^{٤-س} = \frac{١}{١٦}\;{}^{٤}ق_{س}$$

حيث س = ٠ ، ١ ، ٢ ، ٣ ، ٤

(أ) توزيع الاحتمال المطلوب هو التوزيع المبين في العمودين الأول والثاني من الجدول (٣ - ١) الآتي :

الجدول (٣ - ١)

توزيع الاحتمال لعدد الذكور في العائلات ذوات الأربعة الأطفال - ح = ٠,٥

عدد الذكور في العائلة	احتمال هذا العدد	العدد المتوقع من العائلات
س	د (س)	... ٢ × د (س)
٠	$\frac{١}{١٦}\;{}^{٤}ق_{٠} = \frac{١}{١٦}$	١٢٥
١	$\frac{١}{١٦}\;{}^{٤}ق_{١} = \frac{٤}{١٦}$	٥٠٠
٢	$\frac{١}{١٦}\;{}^{٤}ق_{٢} = \frac{٦}{١٦}$	٧٥٠
٣	$\frac{١}{١٦}\;{}^{٤}ق_{٣} = \frac{٤}{١٦}$	٥٠٠
٤	$\frac{١}{١٦}\;{}^{٤}ق_{٤} = \frac{١}{١٦}$	١٢٥
	١,٠٠٠	٢٠٠٠

(ب) أما العمود الثالث فيعطى الأعداد المتوقعة من العائلات التي بها ٠ ، ١ ، ٢ ، ٣ ، ٤ أولاد من بين الـ ٢٠٠٠ عائلة ، ومن هذا العمود ينتج أن العدد المتوقع للعائلات التي بها ولدين على الأقل .

= العدد المتوقع للعائلات التي بها ولدين أو ثلاثة أو أربعة .

= ٧٥٠ + ٥٠٠ + ١٢٥ = ١٣٧٥ عائلة .

مثال (٣ - ٣):

حسب نظرية مندل للخواص الوراثية ، حين يهجن نوع من النباتات حمراء

الزهور مع نوع ذى قرابة من نباتات بيضاء الزهور تنتج خلفة ٢٥٪ منها حمراء الزهور . نفرض أننا سنقوم بتهجين ٥ أزواج من هذه النباتات نختارها عشوائياً فما احتمال أن يكون من بين خمسة السلالات الناتجة :

(أ) لا توجد نباتات حمراء الزهور .

(ب) يوجد على الأقل ٤ نباتات حمراء الزهور ؟

الحل :

لدينا مجتمع من السلالات مقسم إلى القسمين : حمراء الزهور / بيضاء الزهور . واحتمال الحدث « حمراء الزهور » هو عدد ثابت ح = ٠,٢٥ ومن الطبيعى أن نعتبر أن النواتج مستقلة لأن التهجين يتم بين أزواج مختلفة من النباتات . وعلى ذلك تكون الشروط الثلاثة متوفرة ويكون لدينا متغير عشوائي سـ يعبر عن عدد النباتات حمراء الزهور في العينات العشوائية ذوات الحجم ن = ٥ وهو متغير له توزيع ذى الحدين دليلاه ٥ ، ٠,٢٥ ودالة كتلة احتماله

$$\text{د(س)} = {}^{٥}\text{ق}_{\text{س}}\,(٠,٢٥)^{\text{س}}\,(٠,٧٥)^{٥-\text{س}} \quad \text{حيث س} = ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥$$

(أ)احتمال عدم وجود نباتات حمراء الزهور .

$$\text{ل (س} = ٠) = \text{د}(٠) = {}^{٥}\text{ق}_{٠}\,(٠,٢٥)^{٠}\,(٠,٧٥)^{٥} = (٠,٧٥)^{٥}$$

$$= ٠,٢٣٤$$

(ب) احتمال وجود ٤ نباتات حمراء الزهور على الأقل .

$$= \text{ل (سـ} \geqslant ٤) = \text{د}(٤) + \text{د}(٥) \qquad \text{حدثان متنافيان}$$

$$= {}^{٥}\text{ق}_{٤}\,(٠,٢٥)^{٤}\,(٠,٧٥) + (٠,٢٥)^{٥}$$

$$= ٠,٠١٦$$

(٣ - ٣ - ١) تقدير الدليل ح :

فى توزيع ذى الحدين ، إذا كانت قيمة البارامتر ح مجهولة ، يمكن تقديرها تجريبياً من عينات بحيث تتوفر الشروط الثلاثة سالفة الذكر . فإذا حصلنا على التوزيع التكرارى لعينة حجمها ن ووجدنا أن وسطها الحسابي $\bar{س}$ فإننا نأخذ هذا الوسط كتقدير للوسط الحسابي لتوزيع ذى الحدين وهو كما نعلم يساوى ن ح وبالتالى نقدر البارامترح بالعدد ر حيث :

$$ر = \frac{\bar{س}}{ن} \qquad (٧)$$

مثال (٣ - ٤) :

لتقدير نسبة الحصي الجرانيتية إلى مجتمع الحصي على أحد الشواطىء أخذت من هذا الشاطىء ١٠٠ عينة عشوائية بكل منها ثلاث حصوات وحسب عدد الحصوات الجرانيتية بكل عينة فنتج التوزيع التكرارى الآتي :

عدد الحصوات الجرانيتية في العينة	$س_ر$:	٠	١	٢	٣
عدد العينات	$ك_ر$:	٥٨	٣٣	٧	٢

الوسط الحسابي لهذا التوزيع التكرارى = $\bar{س} = \frac{١}{ن} \sum ك_ر س_ر$

$$= \frac{١}{١٠٠} (٠ + ٣٣ + ١٤ + ٦) = ٠,٥٣$$

إذن تقدير البارامتر ح من العينة هو ر = $\frac{\bar{س}}{ن} = \frac{٠,٥٣}{٣} = ٠,١٧٧$ تقريباً

هذا على فرض أن توزيع عدد الحصي الجرانيتية هو توزيع ذى الحدين دليلاه ن ، ح حيث ن = ٣ ، ولنا أن نقول حينئذ أن هناك حوالى ١٧,٧٪ من الحصي الجرانيتية في مجتمع الحصي الذى على ذلك الشاطىء . ويمكننا اختبار مدى صواب هذا القول كما في البند التالى .

ويلاحظ أنه يمكن أيضاً تقدير النسبة ح من عينة عشوائية واحدة بشرط أن تكون كبيرة الحجم وذلك بأخذ التكرار النسبي لعدد الحصي الجرانيتية التي ظهرت في العينة . ففى هذا المثال لدينا ٣٠٠ حصوة منها ٥٣ حصوة جرانيتية (٠ + ٣٣ + ١٤ + ٦) وعلى ذلك فالتكرار النسبى للحصى الجرانيتية هو $\frac{٥٣}{٣٠٠}$ = ٠,١٧٧

(٣ – ٣ – ٢) اختبار ما إذا كان الدليل ح له قيمة معينة . توفيق توزيع ذى الحدين لتوزيع تكرارى معلوم .

نفرض أن قائلا ذكر أن الدليل ح لتوزيع ذى الحدين لمتغير ما له قيمة معينة أ مثلا ونريد اختبار هذا القول . لتحقيق هذا الغرض نتبع الخطوات الثلاث الآتية :

(أ) نختار عينات عشوائية من حجم معين ن ونحسب عدد مرات وقوع الحدث في كل منها أى نحسب العدد س (حيث س = ٠ ، ١ ، ٢ ، ... ، ن) في كل منها . نجمع هذه البيانات في توزيع تكرارى لنحصل على التكرارات $ك_٠$ ، $ك_١$ ، $ك_٢$ ، ... ، $ك_ن$ المناظرة للأعداد ٠ ، ١ ، ٢ ، ... ، ن . إن هذه التكرارات نرمز لها بالرمز $ك_ر$ ونسميها بالتكرارات المشاهدة . إن التوزيع الناتج يكون على نمط التوزيع الوارد بالمثال (٣ – ٤) .

(ب) إذا كان القول أو الفرض ح = أ صحيحا فإن توزيع الاحتمال للمتغير ذى الحدين يكون دليلاه ن ، أ معروفين ونستطيع إيجاد هذا التوزيع والحصول على الاحتمالات $ل_٠$ ، $ل_١$ ، $ل_٢$ ، ... ، $ل_ن$ المناظرة للأعداد ٠ ، ١ ، ٢ ، ... ، ن . نضرب كلا من هذه الاحتمالات (التكرارات النسبية النظرية أو المتوقعة) في عدد العينات المأخوذة لنحصل على الأعداد $ق_٠$ ، $ق_١$ ، $ق_٢$ ، ... ، $ق_ن$. إن هذه الأعداد نرمز لها بالرمز $ق_ر$ ونسميها بالتكرارات النظرية أو المتوقعة .

(حـ) نقارن بين التكرارات النظرية $ق_ر$ والتكرارات المشاهدة $ك_ر$ المناظرة لها فإذا كانت المطابقة حسنة أى كانت أزواج التكرارات قريبة من بعضها بدرجة

معقولة بحيث لا يكون بينها فروق كبيرة جاز لنا قبول الفرض أن ح = أ وإلا نرفضه .

إن مثل هذا الاختبار يدخل في موضوع اختبارات الفروض الذى سنتناوله في فصل لاحق حيث سنتعرف على أدوات تمكننا من الحكم حكماً موضوعياً على مدى صغر أو كبر مثل هذه الفروق وبالتالى من الحكم على صواب، أو خطأ ذلك الفرض .

مثال (٣ – ٥) :

لاختبار الفرض القائل أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال ، اختيرت عشوائياً ٣٢٠ عائلة بكل منها ٤ أطفال فنتج التوزيع التكرارى الآتي :

عدد الأطفال الذكور في العائلة $س_ر$	:	٠	١	٢	٣	٤
عدد العائلات $ك_ر$	:	١٣	٤٣	٩٢	١١٢	٦٠

هل هذه البيانات تدعم الفرض المذكور أو تنفيه ؟

الحل :

على فرض أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال ، فإن احتمال ولادة مولود ذكر يكون ح = $\frac{١}{٢}$ وإذا كان هذا الفرض صحيحاً يكون توزيع عدد الذكور في العائلات ذوات الأربعة الأطفال هو توزيع ذى الحدين دليلاه ٤ ، $\frac{١}{٢}$. وكما في المثال (٣ – ٢) يتولد لدينا التوزيع الآتي :

عدد الأطفال الذكور في العائلة $س_ر$	:	٠	١	٢	٣	٤
احتمال هذا العدد $ل_ر$	:	$\frac{١}{١٦}$	$\frac{٤}{١٦}$	$\frac{٦}{١٦}$	$\frac{٤}{١٦}$	$\frac{١}{١٦}$

ونظراً لأن عدد العائلات في العينات المأخوذة ٣٢٠ فإن التكرارات النظرية التي يمكن مقارنتها بالتكرارات المشاهدة تنتج بضرب هذه الاحتمالات في ٣٢٠ ونحصل بذلك على التوزيع التكرارى النظرى الآتي :

عدد الأطفال في العائلة $س_ر$	:	٠	١	٢	٣	٤
العدد المتوقع للعائلات $ق_ر$	:	٢٠	٨٠	١٢٠	٨٠	٢٠ (المجموع ٣٢٠)

ونقول هنا أننا قد وفقنا توزيع ذى الحدين للتوزيع التكرارى المعطى .

بمقارنة التكرارات النظرية $ق_ر$ بالتكرارات المشاهدة $ك_ر$ وهى :

$ق_ر$	:	٢٠	٨٠	١٢٠	٨٠	٢٠
$ك_ر$	:	١٣	٤٣	٩٢	١١٢	٦٠ (المجموع ٣٢٠)

نجد أن هناك تفاوتاً كبيراً بينها . وعلى فرض توفر شرط العشوائية والاستقلال فإن هذا التفاوت يعزى إلى خطأ الفرض أن ح = $\frac{1}{2}$.

وينبغى أن نشير هنا مرة أخرى إلى أن حكمنا هذا هو حكم ذاتي قد لا يكون هو الحكم الصحيح ، أما الحكم الموضوعى فيستلزم اللجوء إلى أحد الاختبارات الإحصائية المناسبة مثل اختبار χ^2 الذى سندرسه بعد . انظر المثال (٦ – ٨) في البند (٦ – ٧ – ٢) .

تمارين (٣ – ١)

١ – في نوع من أبصال الزهور المعروف أن معدل الإنبات ٩٥٪ . تعبأ هذه الأبصال وتباع في عبوات يحتوى كل منها على ١٠ أبصال . إذا سحب أحد هذه العبوات عشوائياً وزرع ما بها من أبصال فما احتمال كل من الحدثين الآتيين :
(أ) لا تنبت أى بصلة .
(ب) تنبت بصلة واحدة على الأقل ؟

٢ – معدل الإصابة بمرض ما في نوع من البقر ٢٥٪ اختيرت عينة عشوائية من ٨ بقرات . أوجد :

(أ) احتمال أن تكون بقرتان بالضبط مصابتين .

(ب) الوسط الحسابي والتباين لعدد البقرات المصابة في العينات من الحجم ٨ . لماذا ينبغي أن نفترض هنا أن هذا المرض غير معد للبقر ؟

٣ – احتمال إصابة هدف ثابت ٠,٢ . إذا أطلقت ٥ طلقات مستقلة على هذا الهدف فما احتمال إصابته مرة واحدة على الأقل ؟

٤ – لاختبار الفرض القائل أن نسبة الحصي الجرانيتية إلى مجتمع الحصي على أحد الشواطىء هو ح = ٠,٠٢ . أخذت من هذا الشاطىء ١٠٠ عينة عشوائية بكل منها ثلاث حصوات وحسب عدد الحصوات الجرانيتية بكل عينة فنتج التوزيع التكرارى المدون في المثال (٣ – ٤) وهو :

عدد الحصوات الجرانيتية في العينة :	٠	١	٢	٣
عدد العينات :	٥٨	٣٣	٧	٢

اختبر ما إذا كانت هذه البيانات تدعم الفرض المذكور على فرض أن توزيع عدد الحصي الجرانيتية هو توزيع ذى الحدين .

(٣ – ٤) توزيع بواسون : POISSON DISTRIBUTION

توزيع بواسون هو توزيع احتمال لمتغير عشوائي وثاب ~س تأخذ دالة كتلة احتماله الصورة

$$د(س) = \frac{م^{س}}{س!} هـ^{-م} \quad \text{حيث } س = ٠ ، ١ ، ٢ ، \ldots ، م > ٠ \qquad (٨)$$

وحيث هـ أساس اللوغاريتمات الطبيعية (هـ = ٢,٧١٨٢٨ تقريباً)

ولهذه الدالة دليل واحد هو العدد م وبالتالى يتحدد التوزيع تماماً إذا عرفت قيمة م . فمثلا حين م = ٢ يكون التوزيع كالآتي :

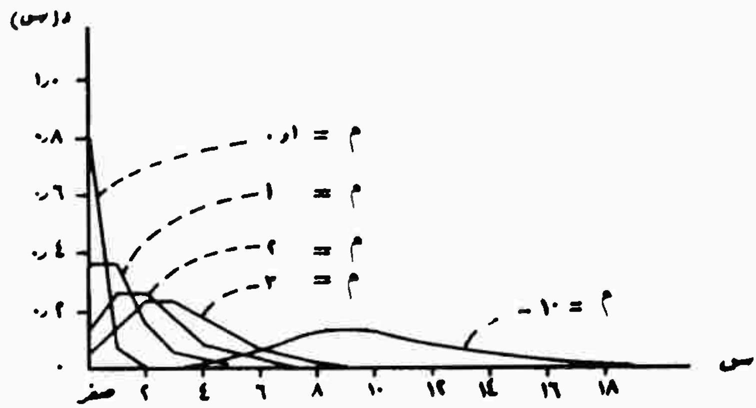
س :	٠	١	٢	٣	٤	٥	٦	...
د (س) :	$هـ^{-٢}$	$٢ هـ^{-٢}$	$٢ هـ^{-٢}$	$\frac{٤}{٣} هـ^{-٢}$	$\frac{٢}{٣} هـ^{-٢}$	$\frac{٤}{١٥} هـ^{-٢}$	$\frac{٤}{٤٥} هـ^{-٢}$	...
أى د(س):	٠,١٣٥	٠,٢٧١	٠,٢٧١	٠,١٨٠	٠,٠٩٠	٠,٠٣٦	٠,٠١٢	...

ومن المميزات الرئيسية لتوزيع بواسون أن

الوسط الحسابي = التباين = م　　　　　　(٩)

ملاحظة (١) :

مجموعة قيم المتغير البواسونى س هى مجموعة لا نهائية {٠ ، ١ ، ٢ ، ...} إلا أنه بعد قيمة معينة س تتوقف على الدليل م ، تتناقص احتمالات هذه القيم تدريجياً حتي تكاد تنعدم كما يشير إلى ذلك الشكل (٣ – ٢) الآتي :



الشكل (٣ – ٢) المضلعات التكرارية لتوزيع بواسون لقيم مختلفة للوسط الحسابي م

ملاحظة (٢) :

هناك جداول تعطى قيم هـ^-م لبعض قيم م – انظر الجدول (٤) بملحق الكتاب – كما أن هناك جداول تعطى الاحتمالات والاحتمالات المتجمعة ل (س = أ) ، ل (س ≥ أ) لتوزيع بواسون لبعض قيم الدليل م والعدد أ – انظر الجدول (٥) بملحق الكتاب .

يستفاد من توزيع بواسون في الموضوعين المقدمين بالبندين الآتيين .

(٣ - ٤ - ١) توزيع بواسون كتقريب لتوزيع ذى الحدين :

نستطيع رياضياً إثبات أنه بوضع م = ح ن في توزيع ذى الحدين المعرف بالدالة (٤) فإنه يؤول إلى توزيع بواسون المعرف بالدالة (٨) حين تقترب ن من اللانهاية وتقترب ح من الصفر .

وهذا يعني من الناحية العملية أنه حين يكون حجم العينة ن كبيراً والاحتمال الثابت ح صغيراً فإن الاحتمالات في توزيع ذى الحدين يمكن إيجادها بالتقريب من الدالة (٨) بدلا من الدالة (٤) مع وضع م = ن ح أى بحيث يكون متوسط توزيع بواسون مساوياً لمتوسط توزيع ذى الحدين . وقد وجد أن هذا التقريب يكون جيداً ، أى يمكن التجاوز عن الخطأ الناشيء عنه إذا كانت :

(ن ≥ ٥٠ ، ن ح ≤ ٥) أو (ح ≤ ٠,١ ، ن ح ≤ ٥)   (١٠)

إن هذا التقريب من شأنه تبسيط حساب احتمالات ذى الحدين إذ أن حسابها من الدالة (٨) أسهل بكثير من حسابها من الدالة (٤) خاصة إذا كانت ن خارج الحدود التي وضعت لها جداول ذى الحدين .

مثال (٣ - ٦) :

إذا كان احتمال أن يحدث رد فعل سيء للشخص الذى يحقن بمصل ما هو ٠,٠٠١ . فاحسب احتمال أن تحدث ٤ حالات ردود فعل سيئة من بين ٢٠٠٠ شخص يحقنون بهذا المصل .

الحل :

توزيع عدد الأشخاص الذين يحدث لهم ردود فعل سيئة هو توزيع ذى الحدين دليلاه ن = ٢٠٠٠ ، ح = ٠,٠٠١ ودالة كتلة احتماله هى :

د (س) = $ق^{٢٠٠٠}_{س}$ (٠,٠٠١)^{س} (٠,٩٩٩)$^{٢٠٠٠-س}$ حيث س = ٠ ، ١ ، ٢ ، ... ، ٢٠٠٠ .

احتمال ٤ حالات ردود فعل سيئة هو :

$$ل\ (س = ٤) = د\ (٤) = {}^{٢٠٠٠}ق_{٤}\ (٠,٠٠١)^{٤}\ (٠,٩٩٩)^{١٩٩٦} = ٠,١٠٥٠$$

من الواضح أن حساب هذا الاحتمال لم يكن سهلا فقد تطلب استخدام اللوغاريتمات والحاسب ، غير أنه يمكننا إيجاد الاحتمال المطلوب تقريبياً من توزيع بواسون نظراً لتوفر أحد شروط التقريب وهو أن ن = ٢٠٠٠ أكبر من ٥٠ ، ن ح = ٢ أصغر من خمسة .

نضع م = ن ح = ٢٠٠٠ × ٠,٠٠١ = ٢

$$د\ (س) = \frac{٢^{س}}{\underline{|س}}\ هـ^{-٢}$$ حيث س = ٠ ، ١ ، ٢ ، ...

$$= \frac{٢^{٤}}{\underline{|٤}}\ هـ^{-٢} = \frac{٢}{٣} \times ٠,١٣٥٣ = ٠,٠٩٠٢$$

وهذا الناتج يمكن إيجاده من الجدول (٥) وهو قريب جداً من الناتج السابق وهو ٠,١٠٥٠ .

(٣ - ٤ - ٢) توزيع بواسون كنموذج لتوزيع الأحداث النادرة :

بصرف النظر عن الدور الذى يلعبه توزيع بواسون كتقريب لتوزيع ذى الحدين فإن له شخصيته واستخداماته الخاصة ، فقد دلت التجربة على أنه يصلح كنموذج احتمال لبعض الأحداث التي تقع عشوائياً عبر الزمان أو المكان .

وعلى سبيل المثال وجد أنه باختيار قيمة مناسبة للدليل م فإن توزيع المتغير س الذى يعبر عن عدد جسيمات ألفا التي تنبعث من مادة مشعة في وحدة زمن مناسبة يمكن أن يعتبر توزيعاً بواسونياً . وبالمثل للمتغيرات التي تعبر (في فترات زمن مناسبة) عن عدد التغيرات الفجائية في الصفات الوراثية - عدد حوادث الطائرات - تتابع طلبات الإغاثة على مراكز الإسعاف - تتابع المكالمات التليفونية على مراكز الهاتف - عدد ولادات ٤ توائم في مدينة - عدد حالات الانفلوانزا

التي ترد إلى مستشفي كبير ... كذلك المتغيرات التي تعبر عن عدد الطحالب في مربع على سفح جبل – عدد الطفيليات على أحد العوائل – عدد البكتريا من نوع معين على طبق بترى Petri plate – عدد الأخطاء المطبعية في صفحة كتاب – الخلل الذى يحدث في جهاز معقد ...

إن مثل هذه الحالات ينظر إليها نمطياً على أنها عملية تولد عدداً من التغيرات أو الأحداث (مثل ظهور جسيم الفا ، طحلب ، بكتريا) في وحدة زمن أو وحدة فراغ مناسبة ، وهذه الوحدة مقسمة إلى عدد كبير جداً من الأجزاء الصغيرة جداً سنسميها **لحظات** instants (سواء كان التقسيم من حيث الزمن أو من حيث الفراغ) وكل من هذه اللحظات يعتبر محاولة يمكن أن يقع فيها الحدث أو لايقع ، وبالتالى فإن هناك إمكانية وقوع الحدث في عدد كبير من المرات .

ويتخذ المتغير توزيعاً بواسونياً إذا توفر الشرطان الآتيان :

(أولا) ندرة الحدث :

يشترط أن يكون معدل وقوع الحدث ، أى متوسط عدد مرات وقوعه في وحدة الزمن أو وحدة الفراغ ، صغيراً بالنسبة لغدد المحاولات التي يمكن أن تسفر عن وقوع الحدث . وهذا ما يجعلنا نصف الحدث بأنه حدث نادر . اعتبر مثلا توزيع المتغير س الذى يعبر عن عدد البكتريا في طبق بترى : إن وحدة الفراغ هنا هى طبق بترى الذى ننظر إليه على أنه مكون من عدد كبير من المساحات الميكروسكوبية (لحظات) كل منها قد يشتمل أو لا يشتمل على بكتريا . وجد بالتجربة أن متوسط عدد البكتريا في الطبق هو عدد متواضع بالرغم من أن هناك بالفعل عدداً لا نهائياً من المحاولات التي يمكن أن تنتج عنها بكتريا ، وعلى هذا فالحدث هو حدث نادر . كذلك اعتبر المتغير الذى يعبر عن عدد الأخطاء في صفحة كتاب جيد الطبع . إن هذه الصفحة تتألف من عدد كبير من الكلمات (لحظات) كل منها قد يقع في كتابتها خطأ أو لا يقع وعلى ذلك فهناك إمكانية

وقوع عدد كبير من الأخطاء ، ولكن نظراً لأن معدل وقوع الخطأ هو عدد صغير جداً نعتبر أن هذا الحدث هو حدث نادر .

واعتبار الحدث نادر هو مسألة نسبية تتطلب أن تكون وحدة الزمن أو وحدة الفراغ كبيرة كبراً كافياً ، فمثلا عندما نعد الطحالب على مربع ما يجب أن يكون هذا المربع كبيراً كبراً كافياً يسمح بنمو عدد وافر من الطحالب مادامت الظروف البيولوجية مهيئة لذلك فلا يجوز مثلا أن تكون مساحة المربع ١ سم^{٢} فقط فهذه المساحة أصغر من جعل الطحالب تتوزع بواسونياً . كذلك في تسجيل حالات الانفلوانزا التي ترد إلى مستشفي كبير لا ينبغى أن تقل وحدة الزمن عن أسبوع مثلا ، لإعطاء الفرصة لورود حالات كافية .

(ثانيا) **استقلال الأحداث** (**عشوائية وقوع الأحداث**) :

يشترط أن يكون وقوع الأحداث **عشوائياً** بمعني أن يكون احتمال وقوع أو عدم وقوع الحدث في أى لحظة مستقلا عن وقوعه أو عدم وقوعه في أى لحظة سابقة أو لاحقه غير متداخلة معها وبذلك لا يتأثر وقوع الحدث إلا بالعوامل العشوائية وحدها . فمثلا وجود طحلب في جزء من مربع ما لا ينبغى أن يزيد أو ينقص من احتمال نمو طحالب أخرى في أى جزء آخر من المربع . كذلك تسجيل حالة انفلوانزا في لحظة ما لا يجب أن يؤثر في احتمال تسجيل حالات تالية .

تحت هذين الشرطين اتضح أنه بتقريب جيد إلى حد كبير أو بالضبط يمكن إيجاد احتمال عدد مرات وقوع الحدث في أى فترة زمنية أو فراغية بواسطة توزيع بواسون دليله :

$$\text{م} = \text{ك ن} \qquad (١١)$$

أى من الدالة $\text{د}(\text{س}) = \frac{(\text{ك ن})^{\text{س}}}{\text{س}!}\,\text{هـ}^{-\text{ك ن}}$ حيث س = ٠ ، ١ ، ٢ ، ....

وحيث د (س) هو احتمال وقوع الحدث س من المرات خلال فترة زمنية أو فراغية ز وحيث ك مقدار ثابت موجب يعبر عن متوسط وقوع الحدث في وحدة الزمن أو الفراغ ، وهذا المقدار يحسب تجريبياً .

مثال (٣ – ٧) :

إذا فرض أن البكتريا من نوع معين تتواجد في الماء بمعدل ٢ بكتريا في السنتيمتر المكعب وأن عدد البكتريا يتوزع توزيعاً بواسونيا فأوجد احتمال أنه في عينة عشوائية من ٢ سم^{٣} من الماء : (أ) لا يوجد بكتريا (ب) يوجد ٣ بكتريا على الأقل .

الحل :

لدينا ك = ٢ بكتريا في السنتيمتر المكعب ، ز = ٢ سم^{٣} .

إذن م = ٢ × ٢ = ٤

، د (س) = $\frac{٤^{س}}{س!}$ هـ$^{-٤}$ حيث س = ٠ ، ١ ، ٢ ، ...

(أ) احتمال عدم وجود بكتريا في ٢ سم^{٢} = د (٠) = هـ$^{-٤}$ = ٠,٠١٨٣٢

(ب) احتمال وجود ٣ بكتريا على الأقل في ٢ سم^{٣} = ل (س $\geq$ ٣) .

= ١ – ل (س $\leq$ ٢) = ١ – [د (٠) + د (١) + د (٢)]

= ١ – (١ + ٤ + ٨) هـ$^{-٤}$ = ١ – ١٣ هـ$^{-٤}$.

= ٠,٧٦١٨٤

مثال (٣ – ٨) :

إذا كان متوسط عدد السيكلوب في لتر من ماء بحيرة هو ٢ فما احتمال وجود ٥ سيكلوب على الأقل في عينة من ٣ لترات من ماء البحيرة ؟

(السيكلوب كائن دقيق يطفو على الماء وتقتات عليه الأسماك) .

الحل :

لدينا ك = ٢ سيكلوب في اللتر ، ز = ٣ لتر

إذن م = ٢ × ٣ = ٦

، $د(س) = \frac{٦^{س}}{س!} هـ^{-٦}$ حيث س = ٠ ، ١ ، ٢ ، ....

احتمال وجود ٥ سيكلوب على الأقل = ل (س ⩾ ٥) = ١ − ل (س ⩽ ٤)

= ١ − ٠,٢٨٥

= ٠,٧١٥

(٣ – ٤ – ٣) اختبار استقلال الأحداث النادرة (أو اختبار العشوائية) :

عند تناول حدث نادر يهمنا في كثير من الأحيان دراسة استقلال الأحداث أى دراسة ما إذا كان وقوع الحدث في لحظة ما يزيد أو ينقص من احتمال وقوعه في لحظة تالية كما هو الحال مثلا في دراسة توزيع سوسة الفاصوليا أو توزيع حشرة على نوع من الذباب . ونظراً لأن الحدث النادر لا يتوزع بواسونيا إلا إذا كانت الأحداث مستقلة فإننا نستطيع الحكم على استقلال الأحداث عن طريق اختبار ما إذا كان التوزيع بواسونيا ، وهذا الاختبار يمكن إجراؤه بتوفيق توزيع بواسون لتوزيع تكرارى مشاهد في تجربة كما فعلنا في حالة ذى الحدين في البند (٢ – ٣ – ١) والمثال (٣ – ٤) . فإذا كانت المطابقة حسنة دل ذلك على أن التوزيع بواسونيا وبالتالى تكون الأحداث مستقلة وتقع عشوائياً ، أما إذا لم تكن المطابقة حسنة فنحكم بعدم عشوائية وقوع الأحداث .

مثال (٣ – ٩) :

أجريت تجربة لاختبار توزيع خلايا الخميرة في ٤٠٠ مربع من جهاز الـ hemacytometer (صندوق لعد الخلايا) ووجد التوزيع التكرارى المبين بالعمودين الأول والثانى من الجدول (٣ – ٢) . بالتأمل في هذين العمودين نلاحظ أمرين هامين هما :

(أ) أن ٧٥ من هذه المربعات أى حوالى ١٩٪ منها لا تشتمل على أى خلية ومعظم المربعات (٥٦٪) تحمل إما خلية واحدة أو خليتين وأن ١٧ مربعاً فقط (أى حوالى ٤٪) تحتوى على ٥ خلايا أو أكثر .

(ب) الوسط الحسابي $\bar{س} = \frac{١}{ن}$ مجـ $ك_{ر}\, س_{ر}$

$$= \frac{١}{٤٠٠} (٧٥ \times ٠ + ١٠٣ \times ١ + ١٢١ \times ٢ \times ... + ١ \times ٩)$$

$$= \frac{٧٢٠}{٤٠٠} = ١,٨$$

الجدول (٣ - ٢)

التكرارات المشاهدة والتكرارات البواسونية المتوقعة لعدد خلايا الخميرة في ٤٠٠ مربع من جهاز الهيماسيتومتر

عدد الخلايا في المربع س	التكرارات المشاهدة ك	التكرارات النسبية المتوقعة ل = د (س)	التكرارات المتوقعة و=ل×٤٠٠	الانحرافات ك - و
٠	٧٥	٠,١٦٥٣	٦٦,١	+
١	١٠٣	٠,٢٩٧٥	١١٩,٠	-
٢	١٢١	٠,٢٦٧٨	١٠٧,١	+
٣	٥٤	٠,١٦٠٧	٦٤,٣	-
٤	٣٠	٠,٠٧٢٣	٢٨,٩	+
٥	١٣	٠,٠٢٦٠	١٠,٤	+
٦	٢	٠,٠٠٧٨	٣,١	-
٧	١ } ١٧	٠,٠٠٢٠	٠,٨ } ١٤,٥	+ } +
٨	٠	٠,٠٠٠٤	٠,٢	-
٩	١	٠,٠٠٠١	٠,٠	-
	٤٠٠	٠,٩٩٩٩	٣٩٩,٩	

أى أن متوسط عدد الخلايا في المربع هو ١,٨ خلية وهذا عدد صغير فعلا بالنسبة لسعة كل مربع وبالنسبة لعدد الخلايا التي يحتمل أن تظهر في أى من المربعات .

من هذا يحق لنا أن نعتبر أن الحدث هو حدث نادر ونتوقع بالتالى أن يكون توزيعه بواسونيا إذا توفر شرط الاستقلال . ولاختبار هذا الشرط نوفق توزيع بواسون للتوزيع التكرارى المشاهد مع تقدير الدليل م لذلك التوزيع من العينة أى نأخذ م = ١,٨ فتكون دالة كتلة الاحتمال :

$$د\,(س) = \frac{١,٨^{س}}{س!}\, هـ^{-١,٨} \quad حيث\ س = ٠ ،\ ١ ،\ ٢ ،\ ....$$

نحسب الاحتمالات د (٠) ، د (١) ، د (٢) ، ... كما في العمود الثالث من الجدول (٣ – ٢) ثم نضرب كلا من هذه الاحتمالات (التكرارات النسبية المتوقعة) في حجم التوزيع التكرارى المشاهد وهو ٤٠٠ فنحصل على التكرارات المتوقعة المبينة بالعمود الرابع .

بمقارنة التكرارات المشاهدة بالتكرارات النظرية نجد أن هناك مطابقة حسنة بين التوزيع التكرارى المشاهد وتوزيع بواسون دليله ١,٨ ولا يوجد نمط واضح لانحرافات التكرارات المشاهدة عن التكرارات المتوقعة لها كما يبدو من العمود الخامس وإن كان الحكم الموضوعى لهذا التطابق لا يتأتي إلا بأحد الاختبارات الإحصائية التي سندرسها بعد . انظر المسألة (٦) في تمارين (٦ – ٢) .

ونستنتج من هذا أن توزيع خلايا الخميرة هو توزيع بواسوني وهذا يتضمن أن الأحداث هنا تقع عشوائياً أى مستقلة عن بعضها .

وهناك اختبار آخر يساعد على بيان ما إذا كان التوزيع بواسونياً دون الالتجاء إلى عملية التوفيق ، ويعتمد هذا الاختبار على الخاصة الهامة التي جاءت في المتساوية (٩) عن تساوى التباين والوسط الحسابي في التوزيعات البواسونية . وحين نتناول

عينة عشوائية من مجتمع بواسوني نتوقع وجود هذا التساوى بالتقريب أى نتوقع أن تكون النسبة بين التباين والوسط الحسابي المحسوبين من العينة قريبة من الواحد الصحيح . في المثال الأخير نجد أن :

$$\text{تباين العينة } ع^{٢} = \frac{١}{ن - ١}\left[\text{مجـ } ك_{ر}\, س_{ر}^{٢} - \frac{(\text{مجـ } ك_{ر}\, س_{ر})^{٢}}{ن} \right]$$

$$= \frac{١}{٣٩٩}\left[(٨١\times١ + \ldots + ٤\times١٢١ + ١\times١٠٣ + ٠\times٧٥) - \frac{٧٢٠^{٢}}{٤٠٠} \right]$$

$$= ١,٩٦٥$$

وهذا العدد قريب من الوسط الحسابي ١,٨ كما أن النسبة بينهما $\frac{١,٩٦٥}{١,٨} = ١,٠٩٢$ قريبة من الواحد وهذا يدعم استنتاجنا السابق بأن توزيع المتغير هو توزيع بواسوني وما يتبع ذلك من عشوائية الأحداث . هذا ويمكن اختبار كبر أو صغر النسبة $ع^{٢} / \bar{س}$ عن الواحد بطريقة موضوعية عن طريق اختبار ت الذى سندرسه بعد . انظر المثال (٦ − ٥) بالبند (٦ − ٦ − ٤) .

نمط التجمع ونمط التنافر :

إذا اتضح أن التكرارات المشاهدة تنحرف عن التكرارات المتوقعة لها بشكل جوهرى أو أن النسبة $ع^{٢} / \bar{س}$ أكبر أو أصغر من الواحد بشكل جوهرى فإن هذا يعني أن الأحداث لا تقع مستقلة عن بعضها بل يؤثر وقوع أو عدم وقوع أحدها في وقوع أو عدم وقوع الأحداث الأخرى . وهذا يدعو إلى التساؤل عما إذا كانت التكرارات المشاهدة تنم عن نمط خاص وعن تفسير ما قد يوجد من أنماط . وهذا التفسير لا يستطيع التحليل الإحصائي وحده القيام به فالمرجع الأول في هذا هو معرفة الباحث بظروف التجربة وطبيعة المتغيرات والعوامل التي تؤثر فيها .

وهناك نوعان رئيسيان من الأنماط هما :

(١) نمط التجمع : CLUMPING

في هذا النمط تكون التكرارات المشاهدة أكبر من التكرارات المتوقعة عند ذيلي التوزيع وأصغر منها عند الوسط كما هو الحال في التجربة الملخصة بالجدول (٣ − ٣) الآتي الذى يعرض توزيع الحلم المائي water mite على ٥٨٩ هاموشة . ويوصف هذا النمط بأنه مُعْدى contagious بمعني أن وقوع حدث (ظهور حلمة مثلا) يرفع من احتمال وقوع أحداث أخرى والعكس بالعكس . وفي هذه الحالة تكون النسبة بين التباين والوسط الحسابي في المجتمع أكبر من الواحد .

(٢) نمط التنافر : REPULSION

في هذا النمط تكون التكرارات المشاهدة أصغر من التكرارات المتوقعة عند الذيلين وأكبر منها عند الوسط كما هو الحال في التجربة الملخصة بالجدول (٣ − ٤) الذى يسجل توزيع السوس على ١١٢ نبات فاصوليا . وهنا يكون وقوع الحدث (خروج سوسة مثلا) عائقاً لوقوع أحداث أخرى فيشتمل التوزيع على عدد قليل من المجموعات المتجانسة وعدد كبير من المجموعات المختلطة . وفي هذه الحالة تكون النسبة بين التباين والوسط الحسابي في المجتمع أصغر من الواحد .

بعد حساب الوسط الحسابي والتباين للتوزيع المشاهد نجد ما يلي :

بالنسبة للتجربة الأولى : ع^{٢} / $\bar{\text{س}}$ = ٠,٩٧٠٩ / ٠,٤٣٤٣ = ٢,٢٣٦

هذه النسبة أكبر جوهرياً من الواحد . انظر البند (٦ − ٦ − ٤) .

بالنسبة للتجربة الثانية : ع^{٢} / $\bar{\text{س}}$ = ٠,٢٦٩ / ٠,٤٦٤٣ = ٠,٥٧٩

هذه النسبة أصغر جوهريا من الواحد . انظر البند (٦ − ٦ − ٤) .

الجدول (٣-٣)

التكرارات المشاهدة والتكرارات البواسونية المتوقعة

لعدد الحلم على ٥٨٩ هاموشه (نمط تجمع)

عدد الحلم على الهاموشة	ك		و		ك – و	
٠	٤٤٢		٣٨٠,٧		+	
١	٩١		١٦٦,١		–	
٢	٢٩		٣٦,٢		–	
٣	١٤		٥,٣		+	
٤	٤		٠,٦		+	
٥	٦	٢٧	٠,١	٦	+	+
٦	٢		٠,٠		+	
٧	٠		٠,٠		٠	
٨	١		٠,٠		+	
	٥٨٩		٥٨٩			

الجدول (٣-٤)

التكرارات المشاهدة والتكرارات البواسونية المتوقعة

لعدد السوس الذى خرج من ١١٢ نبات فاصوليا (نمط تنافر)

عدد السوس على النبات	ك		و		ك – و	
٠	٦١		٧٠,٤		–	
١	٥٠		٣٢,٧		+	
٢	١		٧,٦		–	
٣	٠	١	١,٢	٨,٩	–	–
٤	٩		٠,١		–	
	١١٢		١١٢			

تمارين (٣ – ٢)

(١) في توزيع بواسون دليله م = ٠,٧٢ أوجد :
ل (س = ٠) ، ل (س = ١) ، ل (س = ٢) ، ل (س > ٢) .
(اعتبر أن $هـ^{-٧٢}$ = ٠,٤٨٦٨)

(٢) دلت الخبرة الطويلة على أن السفن تدخل في إحدى الموانى بمعدل ٣ سفن في الساعة . إذا كان توزيع عدد السفن التي تدخل هذا الميناء هو توزيع بواسون فأوجد احتمال أنه في فترة زمنية قدرها دقيقتان (أ) لا تدخل أى سفينة (ب) تدخل سفينتان على الأقل .

(٣) إذا علم أن الجسيمات تنبعث من مصدر مشع بمعدل ٠,٥ جسيماً في الثانية وأن عدد هذه الجسيمات يتوزع بواسونياً فاحسب احتمال انبعاث ٣ جسيمات أو أكثر في فترة زمنية طولها ٦ ثواني .

(٤) الجدول الآتي يعرض توزيع عدد الحلم (المايت المائية) water mite على نوع من الذباب . وفق توزيعاً بواسونياً لهذا التوزيع واستنتج أن وقوع الأحداث (ظهور الحشرات على الذبابة) ليس عشوائياً . أوجد تباين التوزيع التكرارى المشاهد وقارنه بالوسط الحسابي لتدعيم استنتاجك (ستجد أن النسبة بين التباين والوسط الحسابي ٢,٢٢٥) .

عدد الحشرات على الذبابة	٠	١	٢	٣	٤	٥	٦	٧	٨	المجموع
التكرارات المشاهدة	٤٢٢	٩١	٢٩	١٤	٤	٦	٢	٠	١	٥٨٩

(٥) أجب عن نفس السؤال السابق مستخدماً التوزيع التكرارى الآتي الذى يعرض توزيع عدد السوس على قرن فاصوليا (قيس هذا العدد بعدد الثقوب التي نتجت في قرن الفاصوليا أثناء خروج اليرقات منها) .

عدد السوس في القرن :	٠	١	٢	٣	٤	المجموع
التكرارات المشاهدة :	٦١	٥٠	١	٠	٠	١١٢

(٦) يدخل الزبائن في أحد المحلات بمعدل ٣٠ شخصاً في الساعة فإذا كان توزيع عدد الزبائن هو توزيع بواسوني فأوجد احتمال أنه في فترة زمنية قدرها دقيقتان
(أ) لا يدخل أحد (ب) يدخل شخصان على الأقل .

(٣ - ٥) توزيع باسكال : PASCAL DISTRIBUTION

في توزيع ذى الحدين نعتبر أن حجم العينة هو عدد ثابت ن ونعالج متغيراً عشوائياً س يعبر عن عدد مرات وقوع الحدث في ن من المحاولات ، إلا أن هناك مشكلات تتطلب عكس هذا الوضع فيكون حجم العينة متغيراً عشوائياً س ويكون عدد مرات وقوع الحدث هو عدد ثابت أ محدد من قبل ويكون المطلوب هو إيجاد احتمالات قيم المتغير س التي تسمح بوقوع الحدث هذا العدد المحدد من المرات . (يلاحظ أن المعاينة هنا تكون من النوع التتابعى sequential sampling)

أى أننا نعالج متغيراً عشوائياً وثاباً س يعبر عن العدد اللازم من المحاولات لكى يقع الحدث عدداً محددا أ من المرات .

وبطبيعة الحال لا يجب أن يقل عدد المحاولات عن العدد أ لأن وقوع الحدث أ من المرات يتطلب أ محاولة على الأقل ، وعلى ذلك تبدأ قيم س بالعدد أ . أى أن هذا المتغير لا يأخذ إلا القيم أ ، أ + ١ ، أ + ٢ ، ... فإذا كان المطلوب وقوع الحدث ٤ مرات مثلا فإن س تأخذ القيم ٤ ، ٥ ، ٦ ، ...

تحت نفس افتراضات العشوائية وثبات الدليل ح واستقلال الأحداث يمكن أن نثبت رياضياً أن دالة كتلة الاحتمال لهذا المتغير تأخذ الصورة الآتية :

$$\text{د}(\text{س}) = \text{ق}^{\,\text{س}-1}_{\,\text{أ}-1}\ \text{ح}^{\text{أ}}\ \text{ك}^{\text{س}-\text{أ}} \qquad \text{حيث س} = \text{أ ، أ}+1 \text{ ، أ}+2 \text{ ، } \ldots \qquad (12)$$

وحيث ١ > ح >٠ ، ك = ١ - ح ، أ مقدار ثابت

ويسمى توزيع الاحتمال حينئذ بتوزيع باسكال دليلاه ح ، أ كما يسمى المتغير س بمتغير باسكال . يلاحظ أن الحدث يقع أ من المرات المستقله باحتمال ح فى كل مرة وأنه يفشل فى س – أ من المرات المستقلة باحتمال ك فى كل مرة ، وأن المرة الأخيرة هى نجاح دائما . كما يلاحظ أن عدد الطرق التى تقع فيها (أ – ١) نجاحات من (س – ١) محاولات هو $\text{ق}^{\text{س}-1}_{\text{أ}-1}$ وعلى ذلك فإن احتمال وقوع الحدث أ من المرات (وفشله س – أ من المرات) هو $\text{ق}^{\text{س}-1}_{\text{أ}-1}\ \text{ح}^{\text{أ}}\ \text{ك}^{\text{س}-\text{أ}}$

مثال (٣ – ١٠) :

المعروف أن ٦٠٪ من المرضي بمرض معين يستجيبون لدواء ما بعد تناوله لمدة أسبوع . يختار مرضي بهذا المرض الواحد بعد الآخر في ترتيب عشوائي ليتناولوا الدواء (لمدة أسبوع) حتي نحصل على ٥ استجابات صحيحة :

(أ) ما احتمال أن يكون العدد المختار من المرضي سبعة ؟

(ب) ما احتمال أن يكون العدد المختار من المرضي عشرة ؟

الحل :

إن عدد المرضي اللازمين للحصول على خمس استجابات صحيحة هو متغير عشوائي س له توزيع باسكال دليلاه ٠,٦ ، ٥ ودالة كتلة احتماله :

$$\text{د}(\text{س}) = \text{ق}^{\text{س}-1}_{4}\ (0{,}6)^{5}\ (0{,}4)^{\text{س}-5} \qquad \text{حيث س} = 5 ، 6 ، 7 ، \ldots$$

$$\text{(أ) ل } (\text{س}=7) = \text{د}(7) = \text{ق}^{6}_{4}\ (0{,}6)^{5}\ (0{,}4)^{2}$$

$$= 0{,}1866$$

$$\text{(ب) ل } (\text{س}=10) = \text{د}(10) = \text{ق}^{9}_{4}\ (0{,}6)^{5}\ (0{,}4)^{5}$$

$$= 0{,}1003$$

ملاحظة :

إن الحالات التى يظهر فيها متغير باسكالى تنشأ فى المعتاد عندما يستخدم ما يسمى بالمعاينة التتابعية sequential sampling حيث لا يحدد حجم العينة مسبقا ، بل تختار المشاهدات فى تتابع عشوائى الواحدة بعد الأخرى وتتوقف هذه العملية حين يتجمع عدد كاف من المشاهدات يمكننا من اتخاذ قرار بحسب قاعدة معينة توضع سلفا . ففى المثال (٣ - ١٠) نفرض أن المطلوب اختبار صحة الفرض أن ٦٠٪ من المرضى يستجيبون للدواء ، وأن القاعدة التى وضعت لتحديد صحة أو خطأ هذا الفرض كالآتى :

« اختر المرضى الواحد بعد الآخر بترتيب عشوائى ليتناول الدواء وسجل العدد سَ لعدد المرضى المختبرين حتى تحصل على ٥ استجابات صحيحة . ارفض الفرض إذا كان الاحتمال ل (س ≥ سَ) يساوى أو يقل عن ٠,٠٥ وإلا فاقبل الفرض » .

إذا استخدمت هذه القاعدة فماذا يكون حكمنا عن الفرض إذا وجد فى تجربة ما أن عدد المرضى المختبرين حتى الوصول إلى ٥ استجابات صحيحة هو :

(أولا) سَ = ٨ (ثانيا) سَ = ١٣

الحل :

(أولاً) ل(س ≥٨) = ١- ل(س <٨) = ١- [د(٥)+ د(٦)+ د(٧)]

=١-[(٠,٦)^٥+٥×(٠,٦)^٥×٠,٤+١٥×(٠,٦)^٥×(٠,٤)^٢]

= ١-(٠,٦)^٥ [١+٢+٢,٤] = ٠,٥٨

بما أن ٠,٥٨ أكبر من ٠,٠٥ نقبل الفرض أن نسبة الشفاء ٦٠٪

(ثانياً) ل(س≥١٣)=١-ل(س<١٣)=١-[د(٥)+د(٦)+...+د(١٢)]

==١-(٠,٦)^٥[١+٢+٢,٤+...+٠,٣٢٤]=٠,٠٣

بما أن ٠,٠٣ أصغر من ٠,٠٥ نرفض الفرض أن نسبة الشفاء ٦٠٪ .

(٣ – ٦) التوزيع الهندسي : THE GEOMETRIC DISTRIBUTION

هو حالة خاصة من توزيع باسكال تكون فيها عدد مرات وقوع الحدث أ = ١ ويرمز المتغير س هنا إلى عدد المحاولات اللازمة لوقوع الحدث **لأول** مرة . وتنتج دالة كتلة الاحتمال لهذا المتغير بوضع أ = ١ في الصيغة (١٢) أى تكون على الصورة :

د (س) = ح ك^س-١ حيث س = ١ ، ٢ ، ٣ ، ... (١٣)

ويسمى التوزيع حينئذ بالتوزيع الهندسي أو بتوزيع وقت الانتظار waiting time distribution ولهذا التوزيع دليل واحد هو ح . وهو يفيد في دراسة الخواص النادرة للمجتمعات كما هو الحال في أمراض الدم النادرة حيث قد لا ينفعنا استخدام توزيع ذى الحدين ، لأننا لو حددنا حجم العينة فقد لا يقع الحدث في أى عنصر منها وبذلك لا نحصل على معلومات كافية عن الحدث ، بينما التوزيع الهندسي يضمن وقوع الحدث .

يمكن أن نثبت رياضياً أن :

الوسط الحسابي للتوزيع = ١ / ح (١٤)

، تباين التوزيع = ك / ح^٢ (١٥)

مثال (٣ – ١١) :

في إحدى المنتجات الصناعية المعروف أنه في المتوسط توجد وحدة معيبة في كل ١٠٠ وحدة . تختار وحدات عشوائياً وتختبر الواحدة بعد الأخرى إلى أن تظهر أول وحدة معيبة . (أ) ما احتمال اختبار ٥ وحدات حتي الوصول إلى الوحدة المعيبة ؟ (ب) ما العدد المتوقع للاختبارات اللازمة للعثور على أول وحدة معيبة ؟

الحل :

لدينا توزيع هندسي دليله ح = ٠,٠١ وإذن دالة كتلة احتماله

د (س) = ح ك$^{\text{س}-1}$ = ٠,٠١ (٠,٩٩)$^{\text{س}-1}$ حيث س = ١ ، ٢ ، ٣ ، ...

(أ) د (٥) = ٠,٠١ (٠,٩٩)^{٤}

= ٠,٠٠٩٦

(ب) العدد المتوقع يعني الوسط الحسابي = $\frac{1}{\text{ح}}$ = $\frac{1}{0,01}$ = ١٠٠ اختبار .

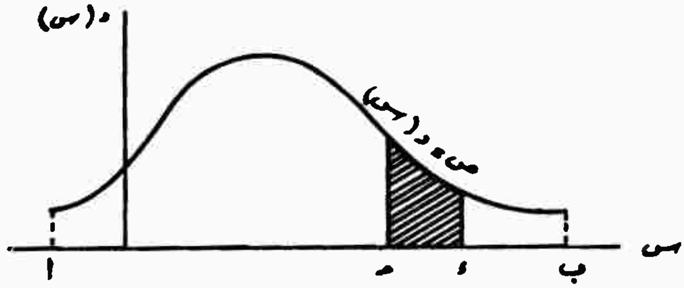
(٣ – ٧) توزيعات الاحتمال المتصلة :

تمتد الأفكار السابقة عن توزيعات الاحتمال للمتغيرات الوثابة إلى حالة المتغيرات المتصلة مع بعض الفروق التي تقتضيها طبيعة كل من هذين النوعين من المتغيرات . فإذا كان س متغيراً حقيقياً من النوع المتصل مداه الفترة (أ ، ب) فإن احتمال أن يأخذ هذا المتغير قيمة محددة س يساوى صفرا ، ذلك لأن أى متغير متصل يأخذ عدداً لا نهائياً من القيم في أى جزء من مداه مهما كان صغيراً .

ولذلك يعرف توزيع الاحتمال في هذه الحالة بواسطة دالة متصلة غير سالبة د حيث :

$$\text{ل}(\text{س}_0+\Delta\text{س} \geq \text{س} > \text{س}_0) = \int_{\text{س}_0}^{\text{س}_0+\Delta\text{س}} \text{د}(\text{س}) .\ \text{ءس} \qquad (١٦)$$

وهذه الصيغة تعني أن احتمال وقوع قيم المتغير س في فترة Δ س يساوى تكامل الدالة د على هذه الفترة . إن مثل هذه الدالة تسمى **بدالة كثافة الاحتمال** probability density function وتمثل بيانا بالشكل (٣ – ٣) .



الشكل (٣ - ٣) توزيع الاحتمال لمتغير متصل

ومن الخواص الرئيسية للمنحني الممثل لأى دالة كثافة احتمال أن مساحة المنطقة الواقعة أسفله وفوق المحور السيني تساوى الواحد الصحيح ، ويمكن رؤية هذا المنحني كخط ممهد لمضلع تكرارى يمثل التكرارات النسبية لتوزيع تكرارى ذى فئات مبني على عدد كبير جداً من المشاهدات موزعة على عدد كبير جداً من الفئات ذوات الأطوال الصغيرة جداً .

ومن تعريف الدالة د نرى أن احتمال وقوع قيم المتغير س بين عددين جـ ، د أى في فترة (جـ ، د) محتواة في المدى (أ ، ب) يعطى بالتكامل .

$$ل(جـ \leq س \leq د) = \int_{جـ}^{د} د(س) \, . \, ءس \qquad (١٧)$$

أى أننا إذا اخترنا عشوائياً قيمة واحدة من قيم المتغير فإن احتمال وقوعها بين عددين جـ ، د يعطى بالتكامل المذكور . ومن الواضح أن هذا التكامل يساوى عددياً مساحة الجزء المظلل بالشكل (٣ ــ ٣) ، كما أنه يعبر عن نسبة قيم المتغير الواقعة بين عددين جـ ، د .

ويعرف الوسط الحسابي μ والتباين σ^2 لتوزيع احتمال متغير متصل مداه الفترة (أ ، ب) كالآتي :

$$\mu = \int_{\text{أ}}^{\text{ب}} \text{س د (س). ءس} \qquad (18)$$

$$،\ \sigma^2 = \int_{\text{أ}}^{\text{ب}} (\text{س} - \mu)^2 \ \text{د (س). ء س} \qquad (19)$$

أو

$$\int_{\text{أ}}^{\text{ب}} \text{س}^2 \ \text{د (س). ء س} - \mu^2$$

مثال (٣ - ١٢) :

أوجد الوسط الحسابي والتباين لتوزيع احتمال المتغير المتصل الذى دالة كثافة احتماله هى :

د (س) = ٦ س (١ - س) حيث ١ > س > ٠ .

الحل :

$$\mu = \int_{0}^{1} 6\,\text{س}^2 (1 - \text{س})\ \text{ء س} = \frac{1}{2}$$

$$\sigma^2 = \int_{0}^{1} 6\,\text{س}^3 (1 - \text{س})\ \text{ءس} - \frac{1}{4} = \frac{1}{20}$$

ومن أشهر توزيعات الاحتمال المتصلة وأكثرها استخداماً ذلك التوزيع المسمى بالتوزيع المعتدل الذى يلعب دوراً رئيسياً في النظرية الإحصائية ونظرية الاحتمال كما أن هناك توزيعات احتمالات متصلة أخرى تستخدم كنماذج احتمال لبعض المجتمعات ، منها التوزيع المعتدل اللوغاريتمى والتوزيع المعتدل الدائرى وتوزيع جاما والتوزيع الأسى ... وسوف نكتفى بتقديم التوزيع المعتدل والتوزيع المعتدل اللوغاريتمى فى الفصل القادم .