الفصل الرابع

نماذج احتمالية لمتغيرات منفصلة

(٤ ـ ١) التجربة الثنائية

يقترن أحد أهم المتغيرات العشوائية المنفصلة بتجربة قذف قطعة نقود، وبالمعنى المجرد للكلمة يُنفذ يوميا العديد من تجارب قذف قطعة النقود ذات الأهمية التطبيقية في العلوم الاجتماعية، والفيزيائية وفي الصناعة وغيرها . . . ، ففي تجارب سبر الرأي العام تشبه مقابلتنا للناخب، من عدة نواح، قذف قطعة نقود. فجوابه «نعم» يوافق وجه الـH ، مثلا، وجوابه «لا» أو امتناعه عن الجواب يقابل الحصول على وجه الـ T .

وهناك أمثلة مشابهة في العلوم الاجتماعية، وفي الصناعة، وفي التربية. إذ يهتم الباحث الاجتماعي بنسبة المنازل الريفية المزودة بالكهرباء. وصانع المنظفات يرغب في تقدير نسبة ربات البيوت اللواتي يفضلن نوعا معينا من المنظفات، ويهتم الأستاذ بتقدير نسبة الطلاب الذين سينجحون في مادته. وسنحصل من كل شخص نقابله على ما يشبه نتيجة قذف قطعة نقود «غير متوازنة بصورة عامة».

والرمي في اتجاه هدف معين يشبه قذف قطعة نقود. فإما أن تكون النتيجة إصابة الهدف، أو عدم إصابته. وإطلاق صاروخ إما أن يكون إطلاقا ناجحا أوفاشلا. وإما أن يكون دواء جديد مفيدا، لمعالجة مرض معين أو لا يكون مفيدا. وإذا اخترنا قطعة مصنعة من خط إنتاج صناعي فإما أن تكون خالية من أي عيب صناعي أو تكون معيبة صناعيا. وتكشف مثل هذه التجارب، على تنوعها، ميزات وخواص التجربة الثنائية.

تعريف التجربة الثنائية

التجربة الثنائية هي تجربة تتصف بالخواص التالية :

١ ـ تتألف التجربة من عدد من التكرارات المتماثلة تماما، n مثلا .

٢ ـ يُنتج كل تكرار إحدى نتيجتين، فإما أن تكون النتيجة «نجاحا»، (أي وقوع الأمر الذي نحن في صدد دراسته) وسنرمز لها بـS، أو أن تكون فشلا، وسنرمز للنتيجة عندئذبـ F .

٣ ـ احتمال النجاح في تكرار معين، وسنرمز له بـ p يبقى ثابتا من تكرار إلى آخر. ويكون احتمال الفشل، بالطبع، $1 - p$ وسنرمز له بـ q.

٤ ـ التكرارات مستقلة بعضها عن بعض

٥ـ نهتم بعدد النجاحات التي نحصل عليها خلال التكرارات الـn ، وسنرمز لهذا العدد بـ X .

وسوف لا تتحقق هذه الشروط جميعها على وجه تام إلا فيما ندر من الحالات العملية. ولكن آثار الحيدان عن هذه الشروط سيبقى بسيطا، ولا يؤثر تأثيرا يُذكر في النتيجة النهائية، طالما بقي هذا الحيدان ضمن حدود معتدلة. فمثلا يبقى احتمال مقابلة ناخب مؤيد للقضية التي ندرسها ثابتا تقريبا من شخص إلى آخر، ما دام مجتمع الناخبين كبيرا جدا بالمقارنة مع العينة من الناخبين الذين تجري مقابلتهم. وإذا كان خمسون بالمائة، مثلا، من مجتمع يحوي ألف ناخب يفضلون المرشح A ، فإن احتمال الحصول على تأييد لـ A في أول مقابلة هو 1/2. واحتمال التأييد في المقابلة الثانية هو 499/999. أو 500/999 ، حسبما تكون المقابلة الأولى قد تمت مع مؤيد أو مع معارض، على الترتيب. والعددان قريبان جدا من 1/2 ، ويمكن اعتبارهما مساويين لـ 1/2 عمليا. كما يمكن الاستمرار في مثل هذا الاعتبار في المقابلة الثالثة والرابعة، وهكذا حتى المقابلة الـ n ، طالما بقي n صغيرا بالنسبة للعدد 1000. وعلى الوجه الآخر، إذا اقتصر المجتمع على عشرة، وكان خمسة منهم يفضلون A، فإن احتمال الحصول على تأييد في المقابلة الأولى هو 1/2 ، ولكنه في الثانية 4/9 أو 5/9، أي أن الاحتمال p يتغير كثيرا من تكرار إلى آخر، ولا يمكن اعتبار التجربة، تجربة ثنائية .

(٤ ـ ٢) دالة التوزيع الثنائي

لنتذكر أولا صيغة نشر ثنائية الحد كما نجدها في كتب الجبر الابتدائية:

$$(q+p)^{n} = q^{n} + \binom{n}{1} p\, q^{n-1} + \binom{n}{2} p^{2}\, q^{n-2} + \ldots + \binom{n}{x} p^{x}\, q^{n-x} + \ldots + p^{n}$$

ونكتب هذا النشر بصورة مختزلة كما يلي:

$$(q+p)^{n} = \sum_{x=0}^{n} \binom{n}{x} p^{x}\, q^{n-x}$$

لنتساءل الآن عن دالة توزيع المتغير العشوائي X، وهو عدد النجاحات الملحوظة في تجربة ثنائية خلال n من التكرارات. والمطلوب ببساطة، وكما رأينا في الفصل السابق، هو الاجابة، بصورة عامة، عن السؤال التالي:

ما هو احتمال أن يأخذ المتغير العشوائي X قيمة تساوي x، أي $P(X = x)$ ؟

وسنجيب عن هذا السؤال في حالة $n = 1$، ثم $n = 2$، ثم $n = 3$، ومن تلمس الخيط المشترك في الحالات الثلاث نحاول استنتاج جواب السؤال المطلوب في الحالة العامة وبالتالي نستنتج صيغة التوزيع.

حالة $n = 1$

الجواب واضح في هذه الحالة من خواص التجربة الثنائية مباشرة، فقيمة X إما أن تكون مساوية للصفر (أي لنتيجة F) أو مساوية للواحد (أي النتيجة S). ونعلم بالفرض أن $P(F) = q$، وأن $P(S) = p$. ومنه جدول التوزيع الاحتمالي المطلوب:

جدول (٤ ـ ١) التوزيع الثنائي النقطي أو توزيع بيرنوللي

X	الاحتمال	نقاط العينة
0	q	F
1	p	S

(أ)

$f(x)$	X
q	0
p	1

(ب)

ويسمى التوزيع في هذه الحالة التوزيع الثنائي النقطي (أو توزيع بيرنوللي). ونلاحظ أن احتمال أن يأخذ X القيمة 0 هو الحد الذي يتضمن p مرفوعا إلى القوة 0 في عبارة $(p+q)^1 = p+q$. (أي الحد الذي لا يظهر فيه الحرف p). وأن احتمال أن يأخذ X القيمة 1 هو الحد الذي يتضمن p مرفوعا إلى القوة 1.

(٤ ـ ٢أ) حالة $n = 2$

في هذه الحالة يبين الجدول (٤ ـ ٢) فضاء العينة والاحتمال الموافق لكل نقطة عينة وقيمة X عند هذه النقطة. ويبين الجدول (٤ ـ ٢) ب دالة التوزيع الاحتمالي لـ X. والقيم الممكنة لـ X هي الآن 0، 1، 2 .

جدول (٤ ـ ٢) التوزيع الثنائي في حالة $n = 2$

نقاط العينة	الاحتمال	X
FF	q^2	0
SF	pq	1
FS	qp	1
SS	p^2	2

(أ)

$f(x)$	X
q^2	0
$2pq$	1
p^2	2

(ب)

وباستعراض العبارة الناتجة عن نشر $(q+p)^2$، وهي

$$q^2 + 2pq + p^2$$

نلاحظ أيضا أن احتمال أن يكون X مساويا للصفر، أي $f(0)$، هو الحد الذي يحوي p مرفوعا إلى القوة صفر (لا تظهر فيه p)، وأن احتمال أن يكون X مساويا للواحد، أي $f(1)$، هو الحد الذي يتضمن p مرفوعا إلى القوة 1، وأن احتمال أن يكون X مساويا للقيمة 2، أي $f(2)$، هو الحد الذي يتضمن p مرفوعا إلى القوة 2.

(٤ ـ ٢ ب) حالة $n = 3$

يتضمن فضاء العينة في هذه الحالة $2^3 = 8$ نقاط، والقيم الممكنة لـ X (عدد النجاحات) هي 0، 1، 2، 3. ويبين الجدول (٤ ـ ٣)أ و(٤ ـ ٣)ب فضاء العينة، ودالة التوزيع، على الترتيب.

جدول (٤ ـ ٣) التوزيع الثنائي في حالة $n = 3$

X	الاحتمال	نقاط العينة
0	q^3	FFF
1	$p\,q^2$	SFF
1	$p\,q^2$	FSF
1	$p\,q^2$	FFS
2	$p^2\,q$	FSS
2	$p^2\,q$	SFS
2	$p^2\,q$	SSF
3	p^3	SSS

(أ)

X	$f(x)$
0	q^3
1	$3p\,q^2$
2	$3\,p^2\,q$
3	p^3

(ب)

ونلاحظ هنا أيضا انطباق القـاعدة التي وجدناهـا في الحالتين السابقتين. فمن الجدول (٤ ـ ٣) ب نجد أن $f\,(0)$ هو الحد الذي يحوي p مرفوعا إلى القوة صفر في نشر ثنائية الحد

$$(q + p)^3 = q^3 + 3pq^2 + 3p^2q + p^3$$

وأن $f\,(1)$ هو الحد الذي يتضمن p مرفوعا إلى القوة 1، وأن $f\,(2)$ هو الحد الذي يحوي p مرفوعا إلى القـوة 2، وأن $f\,(3)$ هو الحـد الذي يحـوي p مرفوعا إلى القوة 3.

وبتعميم هـذه القـاعدة نقول، بصورة عامة، أي في حالة n من التكرارات، إن $f\,(x) = P\,(X = x)$ هـو الحـد الـذي يتضمـن p مرفوعـا إلى القـوة x عنـد نشر ثنائية الحد $(q + p)^n$. ومن صيغة النشر التي استعرضناها في مستهل هذه الفقرة نكتب:

$$f(x) = \binom{n}{x}\, p^x q^{n-x}\ ;\ x = 0, 1, \ldots, n$$

وهي الصيغة العامة لدالة الاحتمال في حالة التوزيع الثنائي.

*وفيما يلي سنقدم عرضا سريعا لاشتقاق رياضي لهذه الصيغة العامة. فما نبغيه هو الاجابة عن السؤال التالي: ما هو احتمال الحصول على x نجاحا عند تكرار التجربة الثنائية n مرة؟ وللاجابة نقول إن احتمال هذه الحادثة، ولنرمز لها بـB، مثلا، هو مجموع احتمالات نقاط العينة التي تنتمي إلى B. وكل نقطة عينة هي متتابعة من n من الحروف F و S، ولكي تنتمي إلى B يجب أن تحوي الحرف S عددا من المرات يساوي x، وتحوي الحرف F عددا من المرات يساوي $n - x$، أي أن لكل نقطة من B الاحتمال نفسه وهو جداء يحوي x مرة العدد p و $n - x$ مرة العدد q. وهكذا يكون احتمال كل نقطة من نقاط الحادثة B مساويا لـ$p^x q^{n-x}$. ويبقى أن نعرف عدد مثل هذه النقاط التي تتضمنها الحادثة B. ولكن هذا العدد ليس إلا عدد إمكانات تقسيم n موضعا إلى زمرتين، تتضمن إحداهما x موضعا، وتتضمن الأخرى $n - x$ موضعا، وبحيث يظهر الحرف S في مواضع الزمرة الأولى ويظهر في مواضع الزمرة الثانية الحرف F. ونعلم أن هذا العدد هو متوافقات n شيئا مأخوذ x منها في وقت واحد، أي $\binom{n}{x}$. ويكون احتمال الحادثة B هو:

$$f(x) = \binom{n}{x} p^x q^{n-x} \; ; \; x = 0, 1, \ldots, n.$$

وتجدر ملاحظة أن

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (q+p)^n = 1^n = 1$$

كما ينبغي أن يكون.

مثال (٤ ـ ١)

لوحظ لفترة طويلة أن صيادا يصيب هدفه باحتمال 0.8. إذا أطلق 4 طلقات على هدف، فما احتمال:

ا ـ إصابة الهدف مرتين؟

ب ـ إصابة الهدف مرتين على الأقل؟

---

* للقراءة فقط.

الحـــل

علينـا أولا تعريف المقصـود بكلمة «نجـاح»، فإذا قلنـا إن النجاح هـو إصابـة الهدف يكون $p = 0.8$، ويصبح $q=1-p=1-0.8=0.2$. والخطوة الثانية هي معرفة عدد التكرارات n، ومن الواضـح هنا أن $n=4$. وبعد تحديد n و p تصبح صيغـة دالة الاحتمال للتوزيع الثنائي محددة تماما. وللإجابـة عن أي سؤال نعبر عنه أولا بدلالة عدد النجاحات X، ثم نطبق صيغة التـوزيع الثنائي لحساب الاحتمال المطلوب. وفي مثـالنا نجد أن صيغة التوزيع الثنائي هي:

$$f(x) = \binom{4}{x}(0.8)^{x}(0.2)^{4-x} \; ; \; x = 0, 1, 2, 3, 4 \, .$$

ا ـ المطلوب $P(X=2)$ أي $f(2)$. وبتعويض x بـ 2 في صيغة التوزيع نجد:

$$f(2) = \binom{4}{2}(0.8)^{2}(0.2)^{2}$$

$$= \frac{4!}{2!\;2!}\,(0.64)(0.04) = 0.1536 \, .$$

ب ـ المطلوب هو $P(X \geq 2)$ ولكن:

$$P(X \geq 2) = P(X=2) + P(X=3) + P(X=4)$$

$$= f(2) + f(3) + f(4)$$

$$= 0.1536 + \binom{4}{3}(0.8)^{3}\,(0.2) + \binom{4}{4}(0.8)^{4}$$

$$= 0.9728$$

والجدير بـالملاحظة أن هذه الاحتمالات سـوف لا تكون صحيحة إذا قـام الرامي بتفقد موقـع الطلقة في كل مرة. ذلك لأنه سيستفيـد من ملاحظاته في الطلقـة التالية، وعنـدها سيكـون من المتـوقع ازدياد قيمـة p من محاولـة إلى أخـرى، وسـوف لا تكـون التكرارات مستقلة كما يقتضي تعريف التجربة الثنائية.

مثال (٤ ـ ٢)

يجري تفتيش الشحنات الكبيرة من البضاعة القادمة إلى مؤسسة صناعية بطريقة العينة. لنفترض أن هذه الطريقة تتلخص في اختيار عشر قطع عشوائيا، ثم اختبارها واحدة فأخرى. وتُرفض البضاعة إذا لاحظنا قطعتين مرفوضتين أو أكثر.

إذا احتوت شحنة بضاعة على 5% من القطع المرفوضة فما هو احتمال قبول البضاعة؟ رفضها؟

الحل

إذا عرفنا النجاح بأنه الحصول على قطعة مرفوضة يكون $p = 0.05$, $q = 0.95$. ومن الواضح أن $n = 10$. وصيغة دالة الاحتمال في مثالنا هي :

$$f(x) = \binom{10}{x}(0.05)^x(0.95)^{10-x} \; ; \; x = 1, 2, \ldots, 10 \; .$$

نعبر الآن عن السؤال المطروح بدلالة عدد النجاحات X فنجد :

P(قبول البضاعة) $= P(X \le 1) = P$ (X يساوي 0 أو 1) $=$

$$P(X=0) + P(X=1) = f(0) + f(1)$$

$$= (0.95)^{10} + \binom{10}{1}(0.05)(0.95)^9 = 0.914$$

P(رفض البضاعة) $= 1 - P$(قبول البضاعة) $= 1 - 0.914 = 0.086$

مثال (٤ ـ ٣)

اختبر لقاح جديد لتحديد فعاليته في الوقاية من الزكام. وقد أعطي لعشرة أشخاص روقبوا لفترة سنة. ووجد أن ثمانية منهم لم يصابوا بالزكام. إذا كان احتمال عدم الإصابة بالزكام خلال سنة هو، بصورة طبيعية، 0.5، فما احتمال ألا يصاب ثمانية أو أكثر علما أن اللقاح لا يزيد في مقاومة الجسم للبرد؟

الحـــل

لنعـرف النجـاح بأنـــه عـــدم الاصـــابة بالزكـام خلال سنـة، فيكون $p = 0.5$، $q = 0.5$، $n = 10$، وتكون دالة الاحتمال لعدد الناجين من الاصابة، X ، هي :

$$f(x) = \binom{10}{x} (0.5)\ (0.5)^{10-x}\ ;\ x = 0, 1, 2, \ldots, 10$$

والمطلوب هو $P(X \geq 8)$، ولحسابه نكتب :

$$\begin{aligned} P(X \geq 8) &= P(X=8) + P(X=9) + P(X=10) \\ &= f(8) + f(9) + f(10) \\ &= \binom{10}{8}(0.5)^8(0.5)^2 + \binom{10}{9}(0.5)^{10} + \binom{10}{10}(0.5)^{10} \\ &= 0.055 \end{aligned}$$

مثال (٤ ـ ٤)

إذا كان 90% من طـلاب مقرر الاحصـاء ينجحـون، فما احتمال فشل اثنين على الأقل من فصل يتضمن عشرين طالبا؟

الحـــل

لنعرف «النجاح» بأنه فشل الطـالب في المقرر. فعندئذ يكون $p = 0.1$ و$q = 0.9$، و$n = 20$. وتكون دالة الاحتمال لعدد الفاشلين، X، هي :

$$f(x) = \binom{20}{x} (0.1)^x (0.9)^{20-x}\ ;\ x = 0, 1, \ldots, 20.$$

أما المطلوب فهو حساب $P(X \geq 2)$. ولدينا :

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) = 1 - P(X=0) - P(X=1) = 1 - f(0) - f(1) \\ &= 1 - (0.9)^{20} - \binom{20}{1}(0.1)(0.9)^{19} \end{aligned}$$

لاحـــظ أن تعـريفنـا للنجاح هنـا كان يتوخى التعبير بسهولـة عن الاحتمال المطلوب. ولو أننـــا عرفنـــا «النجـــاح» بـأنـه نجاح الطالب في المقرر لأصبح $p = 0.9$، $q = 0.1$، $n = 20$، ولأصبحت دالة التوزيع لعدد الناجحين في المقرر، X، هي :

$$f(x)=\binom{20}{x}(0.9)^{x}(0.1)^{20-x}\ ;\ x=0,1,\ldots,20.$$

ويكون المطلوب هو $P\ (X\leq 18)$ لأن فشل اثنين على الأقل يعني أو يكافىء نجاح ثمانية عشر على الأكثر. ولكن

$$P(X\leq 18)=1-P(X\geq 19)=1-P(X=19)-P(X=20)$$
$$=1-f(19)-f(20)=1-\binom{20}{19}(0.9)^{19}(0.1)-\binom{20}{20}(0.9)^{20}$$

وهو الجواب الذي حصلنا عليه سابقا بالضبط.

ونلاحظ من الأمثلة السابقة أن دالة التوزيع الثنائي تقدم علاقة بسيطة لحساب احتمالات حوادث عددية، وهي قابلة للتطبيق في صف واسع من التجارب التي نواجهها في الحياة اليومية. ولكن لابد من الحذر عند استخدام دالة التوزيع الثنائي والتأكد من أن الحالة المدروسة تحقق بصورة مقبولة شروط التجربة الثنائية المذكورة في الفقرة (٤ ـ ١).

وتجدر أيضا ملاحظة أن الأمثلة الأربعة السابقة هي مسائل احتمالية أكثر منها إحصائية. فقد فرضنا أن احتمال النجاح p، وهو الذي يحدد تركيبة المجتمع المدروس، معروف، وكان المطلوب حساب احتمال الحصول على عينة من هذا المجتمع، لها مواصفات محددة. ولو عكسنا الطريقة وافترضنا أننا نملك عينة من مجتمع لا نعرفه ونريد القيام باستقراء حول قيمة p، فعندئذ يقدم المثالان (٤ ـ ٢) و(٤ ـ ٣) مسائل عملية ممتازة يكون الهدف النهائي فيها هو الوصول إلى استقراء إحصائي. وسنناقش هاتين المسألتين بتفصيل أكبر في فقرات قادمة.

(٤ ـ ٣) متوسط التوزيع الثنائي وتباينه*

وفقا لتعريف المتوسط والتباين كما ذكرناهما في الفصل السابق نكتب:

$$\mu=E(X)=\sum_{x=0}^{n}x\cdot f(x)=\sum_{x=0}^{n}x\cdot\frac{n!}{x!\,(n-x)!}\,p^{x}q^{n-x}$$

* البراهين الرياضية للقراءة فقط.

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!\,(n-x)!}\; p^{x}\, q^{n-x}$$

فقد أغفلنا القيمة 0 للمتغير X لأنها ستؤدي عند تعويضها في الحد العام إلى مقدار يساوي الصفر $(0 \times f(x) = 0)$ لنفرض الآن أن $y = x - 1$ فيمكن كتابة العلاقة السابقة على الشكل:

$$\mu = E(X) = \sum_{y=0}^{n-1} \frac{(n-1)!\; n}{y!\,(n-1-y)!}\; p^{y+1}\, q^{(n-1)-y}$$

$$= np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!\,[(n-1)-y]!}\; p^{y}\, q^{(n-1)-y}$$

$$= np\,(p+q)^{n-1} = np.$$

وبطريقة مشابهة نحسب $E[X(X-1)]$ فنجد:

$$E[X(X-1)] = \sum_{x=0}^{n} \frac{x(x-1)\; n!}{x(x-1)(x-2)!\,(n-x)!}\; p^{x}\, q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!\,(n-x)!}\; p^{x}\, q^{n-x}$$

وبوضع $Y = X - 2$، أي $X = Y + 2$ ، نكتب:

$$E[X(X-1)] = n(n-1)\; p^{2} \sum_{y=0}^{n-2} \frac{(n-2)!}{y!\,(n-2-y)!}\; p^{y}\, q^{(n-2)-y}$$

$$= n(n-1)\; p^{2}\,(p+q)^{n-2} = n(n-1)\; p^{2}$$

ولكن:

$$E[X(X-1)] = E(X^{2} - X) = E(X^{2}) - E(X)$$

ومنـه:

$$E(\quad) = E[X(X-1)] + E(X) = n(n-1)\,p^{2} + np$$

وهكذا يكون التباين:

$$\sigma^{2} = E(X-p)^{2} = E(X^{2}) - [E(X)]^{2}$$

$$= n(n-1)\,p^{2} + np - n^{2}p^{2}$$

$$= np - np^{2} = np(1-p) = npq$$

وهكذا نستنتج القاعدة التالية :

لحساب متوسط التوزيع الثنائي نضرب عدد التكرارات n باحتمال النجاح p. ولحساب تباين التوزيع الثنائي نضرب عدد التكرارات n باحتمال النجاح p ثم نضرب الناتج باحتمال الفشل q .

مثال (٤ ـ ٥)

قذفنا 400 ربع ريال على منضدة . ما القيمة المتوقعة والانحراف المعياري لعدد القطع التي تُظهر وجه الـH؟

الحـــل

كل ربع ريال يمثل تكرارا لتجربة قذف قطعة نقود (متزنة) . وإذا اعتبرنا ظهور وجه الـ H نجاحا يكون عدد أوجه الـH الظاهرة، X ، متغيرا يتبع التوزيع الثنـــائي حيث $n = 400$ و $p = 1/2$. ويكون

القيمة المتوقعة لـ X، $E(X) = np = 400 \times 1/2 = 200$

تباين X، $\sigma^2_X = npq = 400 \times \frac{1}{2} \times \frac{1}{2} = 100$

والانحراف المعياري لـ X هو الجذر التربيعي للتباين أي

$$\sigma_X = \sqrt{100} = 10$$

وبما أن التكرارات الـn في التوزيع الثنائي ما هي إلا n تكرارا مستقلا لتجربة ثنائية، فيمكن النظر إلى متغير التوزيع الثنائي X على أنه مجموع n من المتغـــيرات المستقلة X_1، X_2، . . . ، X_n ، أي :

$$X = \sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$$

حيث يخضع كل من هذه المتغيرات في الطرف الأيمن للتوزيع الثنائي النقطي، أي يأخذ كل منها القيمة 1 باحتمال p والقيمة صفر باحتمال $q = 1 - p$. ونعلم من خواص التوقع وخواص التباين أن :

$$V(X) = E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

$$V(X) = V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$$

(المتغيرات مستقلة)

لنأخذ الآن X_1 ولنحسب توقعه وتباينه فنجد من الجدول (٤ ـ ٤) ومن تعريف التوقع والتباين:

جدول (٤ ـ ٤) توزيع X_1

x_1	$f(x_1)$
0	q
1	p

$$E(X_1) = 0 \times q + 1 \times p = p$$

$$E(X_1^2) = 0^2 \times q + 1^2 \times p = p$$

$$V(X_1) = p - p^2 = p(1-p) = pq.$$

وكذلك الأمر بالنسبة لـ X_2 وبقية المتغيرات حتى X_n، فتوقع كل منها يساوي p ما دام احتمال النجاح يبقى نفسه من تكرار إلى آخر، وتباين كل منها pq. ومنه يتضح أن:

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np.$$

وأن:

$$V(X) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} pq = npq.$$

وهي النتائج نفسها التي توصلنا إليها بتطبيق مباشر للتعريف.

تمارين (٤ ـ ١)

١) لنفرض أن واحدا من كل عشرة كتب دراسية للمرحلة الجامعية الأولى يصيب نجاحا باهرا. اختارت دار نشر عشرة كتب جديدة لنشرها. فما احتمال:

ا ـ أن ينال واحد منها فقط نجاحا باهرا؟

ب ـ واحد منها على الأقل يصيب نجاحا باهرا؟

جـ ـ اثنان منها على الأقل يصيبان نجاحا باهرا؟

٢) لنفرض أن المحركات الأربعة لطائرة تجارية تعمل مستقلة بعضها عن بعض. وأن احتمال تعطل أي منها والطائرة في الجو هو 0.1 احسب احتمال:

ا ـ ألا يقع أي عطل والطائرة في الجو.

ب ـ ألا يقع أكثر من عطل واحد.

٣) احتمال كشف جهاز رادار لطائرة معادية هو 0.9. إذا كان لدينا خمسة أجهزة، تعمل مستقلة بعضها عن بعض، فاحسب احتمال:

ا ـ ظهور طائرة معادية على شاشات أربعة منها.

ب ـ اكتشاف وجود طائرة معادية في سمائنا.

٤) قذفنا قطعة نقود متوازنة ثلاث مرات. ليكن X عدد أوجه الـH الملحوظة،

ا ـ اكتب دالة الاحتمال لـX، وارسم مدرجها الاحتمالي.

ب ـ احسب متوسط X وانحرافه المعياري.

جـ ـ بالاستفادة من المدرج الاحتمالي، أوجد النسبة من مجتمع القياسات الواقعة ضمن انحراف معياري واحد على جانبي المتوسط، أعد من أجل انحرافين معياريين. هل تتفق نتائجك مع متباينة تشيبيشيف؟

٥) لنفرض أن قطعة النقود غير متوازنة إلى حد كبير وأن احتمال ظهور وجه الـH هو $p = 0.1$. أعد الخطوتين أ و ب في التمرين السابق ولاحظ كيف تفقد دالة الاحتمال تناظرها عندما لا يكون p مساويا للنصف.

٦) ما احتمال أن يكون أربعة على الأقل من أول ستة أشخاص تقابلهم في الشارع في يوم معين قد وُلدوا يوم الجمعة ($7^6 = 117,649$).

٧) إذا أمكن الافتراض أن عدد المواليد الذكور مساو تقريبا لعدد المواليد الاناث في مجتمع سكاني معين. فأوجد النسبة من الأسر ذوات الستة أطفال التي تتصف بما يلي:

ا ـ عدد الأطفال الذكور يساوي عدد الأطفال الاناث .
ب ـ جميع الأطفال الستة من الجنس نفسه .

٨)ارسم المدرج الاحتمالي للتوزيع الثنائي في كل من الحالات التالية :

أ ـ $p = 0.3$ ، $n = 8$ ؛ ب ـ $p = 0.5$ ، $n = 1$ ؛

جـ ـ $p = 0.1$ ، n=5؛ د ـ $p = 0.1$ ، $n = 10$.

٩) إذا كان 10% من نوع معين من صمامـات التلفزيون يحترق قبل انتهاء مـدة الكفالة . وبيع ألف صمام، فما متـوسط وتبـاين X، حيث X عدد الصمامـات المحترقة قبل انتهـاء مدة كفـالتهـا؟ وما الحدود التي تتـوقع أن يقع xضمنها؟ (استخدم متباينة تشبيشيف) .

١٠) يتضمن جانب من امتحان معين 14 سؤالا من النوع متعدد الاختيارات ، وأمام كل سؤال أربعة أجوبة مقترحة ، واحد منها فقط هو الجواب الصحيح .

ا ـ إذا خصص لكـل سؤال درجة واحـدة فما هي الدرجـة المتوقعة لطـالب يجيب معتمدا على الحزر (أي يختار جوابه عشوائيا)؟

ب ـ إذا خصص لكل إجابة خاطئة 1- فكم يجب أن نخصص للإجـابة الصحيحة حتى تكون الدرجة المتوقعة لطالب يجيب بالحزر صفرا؟

١١) في رحلتك الصباحية إلى الجامعـة تضطر إلى اجتياز 12 مجموعة من إشارات المرور تعمل مستقلة بعضها عن بعض. واحتمال أن تكـون أي منها خضراء عند وصولك إليها هو 1/2. إذا وقفت عنـد أقل من 3 مجموعات فستجد وقتـا لتناول فنجان من الشاي قبل بداية المحاضرة ، وإذا وقفت عند أكثر من 8 مجموعات فستصل إلى قاعة المحاضرات متأخرا. وإذا تأخرت أكثر من مرتين عن موعد المحاضرة خلال أسبوع يتضمن 5 محاضرات صباحية من السبت إلى الاربعاء ، فستتلقى إنذارا من أستاذك .

ا ـ ما احتمال أن تستطيع تناول فنجان شاي في صباحي السبت والأحد؟

ب ـ ما احتمال أن تتلقى إنذارا من أستاذك؟

١٢) إذا كان 5% من البيض الوارد إلى محل لتسويق المواد الغذائية مكسورا، واشتريت 10صناديق في كل منها 6 بيضات، فما احتمال ألا يحتوي أي منها بيضتين أو أكثر من البيض المكسور؟ ما هو في المتوسط عدد الصناديق الذي سيتضمن بيضتين أو أكثر من البيض المكسور؟

١٣) في مدينة كبيرة كان عدم الوفاق بين الزوجين سببا لـ 60% من حالات الطلاق، أوجد احتمال أن ثلاثا من بين حالات الطلاق الست القادمة في هذه المدينة سُيعزى إلى هذا السبب؟

١٤) إذا كان احتمال أن يحتاج طالب إلى وقت إضافي في اختبار الإحصاء هو 0.1، فأوجد احتمال أن اثنين على الأكثر من 5 طلاب سيحتاجون إلى وقت إضافي. ما احتمال ألا يحتاج واحد على الأقل من الطلاب الخمسة إلى وقت إضافي؟

١٥) إذا كان احتمال تحمل نوع من المصابيح للضغط العالي هو 0.4، وأخذنا عينة من 100 مصباح، فما احتمال ألا يتحمل 65 منها الضغط العالي. (أعط صيغة الجواب دون إجراء الحسابات).

١٦) إذا كان 10% من إنتاج آلة معينة يتضمن عيبا صناعيا، وأخذنا عينة عشوائية من 100قطعة، فما احتمال أن يكون ثلاثون منها على الأكثر معيبا؟ (أعط صيغة الجواب دون إجراء الحسابات).

١٧) من سجلات المواليد في إحدى مدن ولاية يوتا الأمريكية تجد فيما يلي بيانا إحصائيا يُظهر جنس كل من الأطفال الأربعة مرتبة حسب تعاقب ولادتهم في كل من 7745 أسرة من الأسر ذات الأربعة أطفال. (M ترمز لذكر F ترمز لأنثى) والمطلوب:

ا ـ استخدام هذه الإحصائية لتقدير نموذج احتمالي مناسب لمجتمع الأسر من أربعة أطفال. (اعتبر أن التجربة هي أن تختار عشوائيا أسرة من مجتمع الأسر ذات الأربعة أطفال وتسجل جنس الأطفال الأربعة حسب تعاقب ولادتهم).

ب ـ استخدام هذه الإحصائية لتقدير التوزيع الاحتمالي لـ X، عدد الأطفال الذكور في أسرة من أسر هذا المجتمع.

جـ ـ استخدام هذه الاحصائية لتقدير احتمال ولادة طفل ذكر في هذا المجتمع. واعتباره احتمال النجاح p لتجربة ثنائية فيها $n = 4$ ، و X عدد الذكور من بين الأربعة.

د ـ اكتب التوزيع الاحتمالي لـ X في السؤال جـ وهو التوزيع النظري وقارنه بالتوزيع الاحتمالي لـ X الذي استنتجته في ب وهو التوزيع التجريبي. هل تعتقد أن التوزيع الثنائي هو النموذج الاحتمالي المناسب لوصف ودراسة عدد الصبيان في أسرة من مجتمع الأسر ذوات الأربعة أطفال؟ أي وصف مجتمع القياسات للمتغير العشوائي X، الذي يرمز إلى عدد الذكور في أسرة من هذا المجتمع؟ (قم بحساباتك لثلاثة أرقام عشرية).

التكرار	جنس الأطفال في الأسرة حسب ترتيب ولادتهم	التكرار	جنس الأطفال في الأسرة حسب ترتيب ولادتهم
526	MFFM	537	MMMM
498	FMFM	549	MMMF
490	FFMM	514	MMFM
429	MFFF	523	MFMM
451	FMFF	467	FMMM
456	FFMF	497	MMFF
441	FFFM	486	MFMF
408	FFFF	473	FMMF

١٨) وجدت شركة طيران أنه، في المتوسط، يفشل 4 بالمائة من المسافرين الذين يحجزون مقاعد لرحلة معينة في الوصول إلى قاعة المسافرين في الوقت المحدد. ولذلك قررت

الشركة السماح لـ 75 شخصا أن يحجزوا مقاعدهم في طائرة لا تتسع إلا لثلاثة وسبعين راكبا. ما احتمال توفر مقعد لكل مسافر يصل في الوقت المحدد؟

(٤ ـ ٤) الكشف على بضاعة بطريقة العينة*

نعلم أن المؤسسة الصناعية هي مكان تتحول فيه مادة أو مواد خام إلى مادة مصنعة. ولابد لإدارة المؤسسة، حفاظا منها على مستوى معين لجودة المصنوعات، أن تجعل كمية المادة الخام غير الصالحة التي تدخل في عملية الإنتاج أصغر ما يمكن. كما ترغب في خفض عدد القطع المنتجة المعيبة صناعيا إلى أقل حد ممكن أيضا. وفي محاولة لبلوغ هذا الهدف تقيم «غربالا» للبضائع الداخلة في عملية الإنتاج والخارجة منها في محاولة لمنع غير المناسب من العبور في الاتجاهين كليهما.

ولتبسيط المناقشة، لنفرض أن ما يهمنا هو «غربلة» البضاعة الواردة، أي المواد الخام المؤلفة، مثلا، من قطع على شكل صناديق من مادة معينة. فإما أن نقبل شحنة البضاعة الواصلة إلى المصنع، إذا كانت نسبة غير الصالح فيها نسبة مقبولة. وإما أن تكون هذه النسبة عالية فنرفض البضاعة ونردها إلى المموّل.

ويمكن إقامة «الغربال» بعدة طرق. ومن الواضح أن أكمل هذه الطرق هي أن يتم الكشف على كامل البضاعة قطعة فأخرى. وللأسف فإن تكاليف مثل هذا الكشف قد تكون كبيرة إلى الحد الذي يجعلها غير واردة البتة من وجهة النظر الاقتصادية. هذا ناهيك عما يمكن أن يُقبل أو يُرفض خطأً من قبل المفتش، خاصة بعد أن ينال منه الجهد مناله نظرا لضخامة العمل المطلوب. يضاف إلى ذلك أنه قد تكون هذه الطريقة مرفوضة بالنظر إلى طبيعة المادة التي نكشف عليها. فاختبار صلاحية المصباح الصغير (الفلاش) المستخدم في آلة تصوير يؤدي إلى تلفه. واختبار كل البضاعة، في مثل هذه الحالة، يعني ألا يبقى شيء لاستخدامه أو لبيعه.

وطريقة الغربلة الثانية الأقل تكلفة، والتي توفر جهودا كبيرة، هي طريقة العينة الإحصائية. وهي مشابهة للخطة التي ذكرناها في المثال (٤ ـ ٢). وفيها نختار عينة من

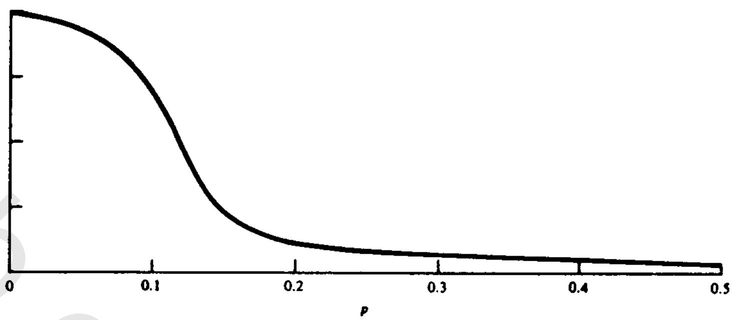
* للقراءة فقط.

n قطعة من قطع الشحنة بطريقة عشوائية، ونكشف عليها بدقة قطعة فأخرى لمعرفة ما تحويه العينة من قطع غير صالحة. وإذا كان عدد هذه القطع، ولنرمز له بـ X، أقل أو يساوي عددا a حددناه سلفا، ويسمى عدد القبول، نقرر قبول البضاعة، وفيما عدا ذلك نرفضها ونعيدها إلى المموّل. وكان عدد القبول في الخطة التي ذكرناها في المثال (٤ ـ ٢) هو $a=1$.

ويلاحظ القارىء أن خطة العينة تعمل بطريقة موضوعية تماما، وتؤدي إلى استقراء يتعلق بمجتمع القطع التي تتألف منها الشحنة. ورفض البضاعة يعني أننا استقرأنا أن نسبة القطع غير المقبولة، p، هي نسبة كبيرة تتجاوز الحد الذي يمكن التساهل فيه، والذي يؤدي إلى تدهور مستوى الجودة في الناتج النهائي في المصنع. وقبول البضاعة يعني أننا استقرأنا أن تلك النسبة، p، صغيرة، وأنها تبقى في حدود المعقول في عملية التصنيع. ويقدم الكشف على بضاعة بطريقة العينة مثالا على عملية اتخاذ قرار إحصائي.

ولا تكون مناقشتنا تامة إذا أهملنا بعض النقاط المتعلقة بجودة الطريقة المستخدمة للقيام بالاستقراء. ومع أن خطة العينة التي عرضناها أعلاه هي طريقة لاتخاذ قرار إلا أنها ليست وحيدة. ويمكننا تغيير حجم العينة n، عدد القبول a، أو اتباع طريقة في اتخاذ القرار غير إحصائية وراجعة للتقديرات الشخصية. فكيف يمكن مقارنة هذه الطر ق المختلفة في اتخاذ القرار؟ والجواب الطبيعي هو أن نختار الطريقة التي تؤدي إلى القرار الصحيح بأكبر تواتر ممكن، أو على الوجه الآخر تؤدي إلى القرار غير الصحيح بأقل نسبة من المرات.

ويميز مهندسو الإنتاج جودة خطة العينة بحساب احتمالات قبول البضاعة في حالة نسب مختلفة للقطع غير الصالحة في الشحنة الواردة. ويمثلون نتائج هذه الحسابات في شكل بياني يدعى «المنحنى العملياتي المميز» لخطة العينة. ويبين الشكل (٤ ـ ١) نموذجا لمثل هذه المنحنيات. ولكي يؤدي الغربال مهمته بصورة مرضية، نرغب في أن يكون احتمال قبول شحنات نسبة العطل فيها ضعيفة مرتفعا، وأن يكون منخفضا في حالة شحنات نسبة العطل فيها مرتفعة. ويلاحظ القارىء أن احتمال القبول سينحدر باستمرار مع ارتفاع نسبة العطل، وهي النتيجة التي نتوقعها.



شكل (٤ ـ ١) نموذج لمنحنى عملياتي مميز لخطة عينة .

وعلى سبيل المثـال، إذا كان الممـوّل مطمئنا إلى أن نسبـة العطل في شحناتـه من البضاعة لا تتجاوز 1%، وكان المصنع يعمل بصورة مرضيـة بشحنات تقل نسبة العطل فيها عن5%، فعندئذ يجب أن يكون احتمال قبول شحنات بنسبة من العطل أقل من 1% مرتفعا . ومـا لم يكن الأمر كذلك فإن المموّل سيرفع أسعاره لتغطية نفقـات إعادة شحنة ممتازة «تحوي أقل من 1% من العطل» إليـه، أو أنه سيحمّل إدارة المصنع نفقـات إعادة الكشف عـلى البضـاعـة . وعلى الوجـه الآخر فإن احتمال قبول شحنـات نسبة العطل فيها 5% أو أكثر لابد أن يكون منخفضا .

مثال (٤ ـ ٦)

احسـب احـتمال قبول شحنـة عند استخـدام خطة عينـة فيهـا حجـم العينة $n = 5$ ، وعـدد القبـول $a = 0$. وذلـك إذا كانـت نسـب القطـع غيـر الصالحـة تساوي $p = 0.1$، $p = 0.3$، $p = 0.5$. ارسم المنحنى العمليـاتي المميـز لهذه الخطة .

الحـل

لدينا توزيع ثنائي فيه $n = 5$، واحتمال النجاح p . وصيغة دالة التوزيع :

$$f(x) = \binom{5}{x} p^x q^{5-x}, \quad x = 0, 1, 2, 3, 4, 5$$

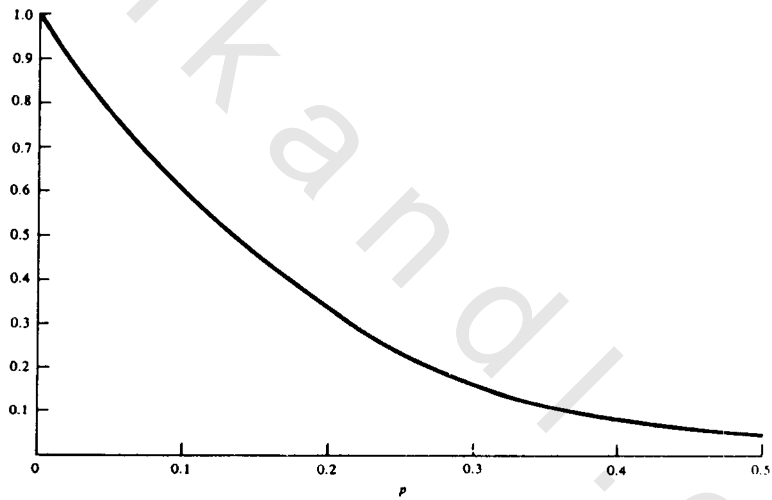
ويكون

$$P(\text{القبول}) = f(0) = q^5$$

ومنه :

$$P(\text{القبول} \mid p = 0.1) = (0.9)^5 = 0.590$$
$$P(\text{القبول} \mid p = 0.3) = (0.7)^5 = 0.160$$
$$P(\text{القبول} \mid p = 0.5) = (0.5)^5 = 0.031$$

ونعلم بالاضافة إلى ذلك أن احتمال القبول يجب أن يكون الواحد عندما يكون $p = 0$، وأن يكون صفرا عندما $p = 1$. وبرسم النقاط الخمسة، حيث الإحداثي السيني هو نسبة العطل. والإحداثي الصادي هو احتمال القبول الموافق، يمكن تخطيط شكل تقريبي للمنحنى العملياتي المميز وهو مبين في الشكل (٤ ـ ٢).



شكل (٤ ـ ٢) لمنحنى العملياتي المميز في حالة $n = 5$ ، $a = 0$

وقد يصبح حساب احتمالات التوزيع الثنائي عملا شاقا في حالة n كبيرة. ولتيسير الحسابات تتوفر عادة جداول تعطي مجموع احتمالات التوزيع الثنائي من $x = 0$ إلى $x = a$ عدد القبول. وذلك في حالة عينات حجمها n يساوي 5، 10، 15، 20، 25 .

مثال (٤ ـ ٧)

ارسم المنحنى العملياتي المميز لخطة عينة فيها $n = 15$ و $a = 1$.

الحـــل

سنحسب احتمال القبول في حالة $p = 0.1$، $p = 0.2$، $p = 0.3$، $p = 0.5$. وهكذا نكتب:

$$\text{احتمال القبول} = \sum_{x=0}^{1} f(x) = f(0) + f(1) = q^{15} + \binom{15}{1} p\, q^{14}$$

ومنه:

$$P(\text{القبول} \mid p = 0.1) = (0.9)^{15} + 15\ (.1)(.9)^{14} = .549$$

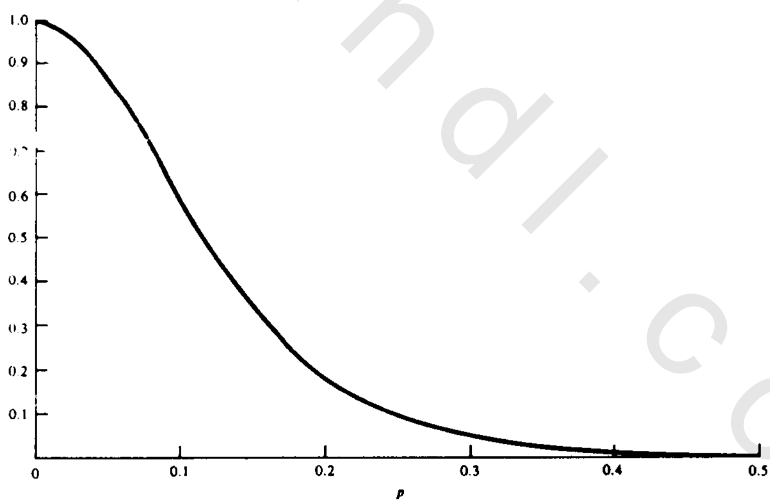
وبصورة مشابهة نجد:

$$P(\text{القبول} \mid p = 0.2) = 0.167$$

$$P(\text{القبول} \mid p = 0.3) = 0.035$$

$$P(\text{القبول} \mid p = 0.5) = 0.000$$

والمنحنى العملياتي المميز مبين في الشكل (٤ ـ ٣).



شكل (٤ ـ ٣): المنحنى العملياتي المميز في حالة $n = 15$ ، $a = 1$

وتستخدم خطة العينة على نطاق واسع في الصناعة. ولكل خطة عينة منحن عملياتي مميز، يميز الخطة عن غيرها، ويقدم نوعا من الوصف لحجم ثقوب الغربال. وسيختار مهندس الإنتاج الخطة بحيث يحقق المتطلبات التي يفرضها واقعه. فزيادة

عدد القبول تزيد من احتمال القبول، وبالتالي توسع ثقوب الغربال. كما تقدم زيادة حجم العينة قدرا أكبر من المعلومات التي يمكن أن نبني عليها قرارنا، وبالتالي تزيد من قدرة الطريقة المتبعة على التمييز. و هكذا ينحدر المنحنى العملياتي المميز بسرعة مع ازدياد p عندما يكون حجم العينة n كبيرا. (قارن بين الشكل (٤ ـ ٢) حيث $n = 5$، والشكل (٤ ـ ٣) حيث $n = 15$).

تمارين (٤ ـ ٢)

١) يتفق شار وبائع على استخدام طريقة الكشف بالعينة مستخدمين عينة حجمها $n = 5$ وعدد قبول $a = 0$. ما هو احتمال أن يقبل الشاري شحنة بضاعة نسبة العطل الحقيقية فيها:
أ ـ $p = 0.1$، ب ـ $p = 0.3$، جـ ـ $p = 0.5$، د ـ $p = 0$، هـ ـ $p = 1$.
ارسم المنحنى العملياتي المميز لهذه الخطة.

٢) أعد التمرين ١ في حالة $n = 5$، $a = 1$.

٣) أعد التمرين ١ في حالة $n = 10$، $a = 0$.

٤) أعد التمرين ١ في حالة $n = 10$، $a = 1$.

٥) ارسم المنحنيات العملياتية المميزة للخطط الأربع في التمارين ١، ٢، ٣، ٤، على ورقة بيانية واحدة. ما تأثير زيادة عدد القبول a مع بقاء n ثابتة؟ وما تأثير زيادة حجم العينة n، عندما تبقى a ثابتة؟

(٤ ـ ٥) اختبار فرضية*

إن مسألة اللقاح ضد الزكام المعطاة في المثال (٤ ـ ٣)، هي مسألة توضيحية لاختبار إحصائي لفرضية. وتتلخص المسألة في السؤال التالي: هل تقدم المعلومات التي تحويها العينة دلالة كافية على فعالية اللقاح؟

* للقراءة فقط.

ويحمل المنطق المستخدم في اختبار فرضية شبهـا كبيرا بالأسلـوب المستخدم في قـاعة محكمـة. فعنـد محاكمة رجل متهم تفترض المحكمـة أن المتهم بـريء حتى تثبت إدانته. ويجمع ممثل النيابة كل الأدلة المتوفرة له ويقدمها في محاولة لنقض فرضية البراءة، وبالتالي الحصول على إدانة المتهم والحكم عليه. وتصـور المسألة الاحصائية اللقاح ضد الزكـام متهما. والفرضية التي سيجـري إختبارها، وتـدعى الفرضية الابتـدائية، هي أن اللقـاح غير فعال. ودلالات الـدعـوى موجـودة ضمن العينـة المسحوبـة من مجتمع. ويعتقد الباحث وهـو يؤدي دور ممثل النيابة أن اللقاح مفيد فعـلا. ويحاول تبعا لذلك استخدام الدلالات المتوافـرة في العينة لرفض الفرضية الابتدائية وبـالتالي دعم قناعته بأن اللقاح، في الحقيقـة، ناجـح جدا ضد الـزكام. (لاحظ أن التهمـة هنا هـي أن اللقاح فعال، والفـرضية الإبتدائية هي بـراءة اللقاح من هذه التهمة) وسيتعـرف القارىء على هذا الأسلوب كشكل أساسي من أشكال الطـريقة العلمية الحديثة حيث يتوجب وضع النظريات المقترحة على محك الواقع.

ويبدو بـديهيا أن نختار عدد من يجانبهم الـزكام، X، كقياس لمقـدار البينة التي تحتويها العينة. وإذا كان X كبيرا فإننا سنميل إلى رفض الفرضيـة الإبتدائية واستنتاج أن اللقاح فعـال. وعلى الوجـه الآخر سيقـدم صغر X القليل من الـدعم لرفض الفـرضية الإبتدائية. وفي الحقيقـة، إذا كانت الفرضية الإبتدائية صحيحـة واللقاح غير فعال فإن احتمال النجاة من الزكام طوال فصل الشتاء سيكون $p = 1/2$، وستكون القيمة المتوسطة لـX هي:

$$E(X) = n\,p = 10\left(\frac{1}{2}\right) = 5\,.$$

وقد لا يجد معظم المهتمين صعوبة في تكوين حكمهم الخاص في حالة $X = 10$ أو في حالة يساوي 5، 4، 3، 2، أو 1 حيث تقدم في الظـاهر دلالة كبيرة لـرفض أو قبول الفرضية، على الترتيب. ولكن ماذا يمكن أن يُقال في حالات أقل وضوحا مثل $X = 7$ أو $X = 8$، أو $X = 9$؟ وسواء اتخذنـا قرارا بطريقـة ذاتية أو موضـوعية، فمن الـواضح أننا سنختار الطريقة التي تعطي أقل احتمال لاتخاذ قرار غير صحيح.

وسيختبر الإحصائي الفرضية الابتدائية بطريقة موضوعية، ولكنها مشابهة لما يمكن أن نصل إليه باللجوء إلى الحس السليم أو الفطرة. وصانع القرار، ويدعى عادة «الإحصاء» يُحسب عادة من العينة. وفي مسألتنا فإن هذا الإحصاء هو عدد من نجوا من الإصابة بالزكام، X، وسنأخذ عندئذ في اعتبارنا كل القيم الممكنة لهذا الإحصاء، وهي هنا، $X = 0, 1, 2, \ldots, 10$ ثم نقسم هذه القيم إلى مجموعتين، ندعو إحداها منطقة الرفض، والأخرى منطقة القبول. وهكذا تُنفذ التجربة، ونلاحظ قيمة «صانع القرار» أو «الإحصاء»، X. فإذا أخذ X قيمة من منطقة الرفض رفضنا الفرضية. وفيما عدا ذلك نقبلها. وعلى سبيل المثال، يمكننا اختيار منطقة الرفض من النقاط $X = 8$ أو 9، أو 10. ونعتبر ما تبقى من قيم X منطقة قبول. وبما أننا لاحظنا القيمة 8 في مثالنا فإننا نرفض الفرضية الإبتدائية بأن اللقاح غير فعال ونستنتج أن احتمال النجاة من الزكام طوال عام كامل هي أكبر من 1/2 عند استخدام اللقاح. والآن ما هو احتمال أن نرفض الفر ضية الإبتدائية مع أنها في الواقع صحيحة؟ واحتمال الرفض الخاطىء للفرضية الإبتدائية هو احتمال أن نأخذ X القيمة 8، 9، أو 10، علما أن $p = 1/2$، وهذا هو، في الحقيقة، الاحتمال الذي حسبناه في المثال (٤ ـ ٣) ووجدناه مساويا لـ 0.055 . وبما أننا قررنا رفض الفرضية الإبتدائية ووجدنا أن احتمال أن يكون هذا الرفض غير صحيح هو احتمال بسيط فإن هذا يولد لدينا ثقة غير قليلة بأننا اتخذنا القرار الصحيح.

عند تأمل المسألة قليلاً سيلاحظ القارىء أن الشركة المنتجة للقاح ستواجه نوعين من الخطأ. فمن جهة يمكن أن ترفض الفرضية الإبتدائية وتستنتج خطأ أن اللقاح فعال. وإنتاج الدفعة الأولى من اللقاح وطرحها للاستخدام سيسبب خسارة مادية ومعنوية (الإساءة إلى سمعة الشركة) لأن الحقيقة ستكشف عن نفسها. ومن الجهة الأخرى، يمكن أن تقرر قبول الفرضية الابتدائية، وتستنتج خطأ أن اللقاح غير فعال. وسيقود هذا الخطأ إلى خسارة الفوائد الجمة الصحية والمادية التي كان سيقدمها طرح ذلك اللقاح المفيد في الأسواق لاستخدامه على نطاق واسع.

ويدعى رفض الفرضية الابتدائية مع أنها صحيحة بالخطأ من النوع الأول (أو النوع I). ونرمز لاحتمال ارتكاب خطأ من النوع الأول بـ α . وسيزداد الاحتمال α أو

يتناقص مع اتساع أو تقلص منطقة الرفض. وبالقدر الذي تمثل فيه α مخاطرة الرفض الخاطىء، يمكن أن نتساءل: لماذا لا نختار منطقة الرفض صغيرة قدر المستطاع، ونقلل بذلك احتمال تلك المخاطرة؟ فمثلا لماذا لا نختار $X = 10$ فقط منطقة رفض في مثالنا هنا؟ ولكن لسوء الحظ إن تخفيض α يزيد من احتمال ارتكاب خطأ من نوع آخر، وهو احتمال قبول الفرضية الابتدائية مع أنها غير صحيحة، وأن الصحيح هو فرضية بديلة تختلف عنها. ويدعى هذا الخطأ بالخطأ من النوع الثاني (أو النوع II). ونرمز لاحتمال مثل هذا الخطأ بالرمز β . أي أن β هو احتمال القبول الخاطىء. ومن أجل حجم ثابت للعينة n، تكون العلاقة بين α و β علاقة عكسية. فعندما يزداد أحدهما يتناقص الآخر. وتقدم زيادة حجم العينة معلومات أكثر يمكن أن نبني عليها قرارنا، وبالتالي تخفض كلا من α و β . ويقيس احتمالا الخطأ من النوعين I و II، أي α و β، مخاطرة التورط بقرار غير صحيح. ويختار المجرب، وفقا لما تمليه طبيعة المسألة المدروسة، حجم هذين الاحتمالين. وعادة نختار حجم العينة n، ونحدد شكل منطقة الرفض وحجمها، بحيث نضع سقفا للاحتمال α لا يتجاوزه، ويسمى مستوى المعنوية، وتحت هذا الشرط نحاول جعل β أصغر ما يمكن. ومن الواضح أن اختيار شكل منطقة الرفض يشكل أمرا حاسما في مسألة الاختبار الإحصائي وتتوقف عليه إلى حد كبير قوة وكفاءة الاختبار الإحصائي.

تمارين (٤ ـ ٣)

١) نقوم بتجربة لاختبار أن قطعة نقود متوازنة، وذلك بقذف قطعة النقود أربع مرات وملاحظة عدد أوجه الصورة التي تظهر. ونرفض الفرضية إذا كان هذا العدد صفرا أو أربعة.

ا ـ ما هو احتمال الخطأ من النوع الثاني؟

ب ـ إذا كانت القطعة فعلا غير متوازنة واحتمال ظهور وجه الصورة هو 0.7، فما هو احتمال الخطأ من النوع الثاني في هذا الاختبار؟

٢) نتوقع أن يعطي زوج من الخنافس نسلا بعينين سوداوين بنسبة 30% من المرات. ولاختبار هذه النظرية نلاحظ ثلاثا من نسلها فنجد أن عيونها زرقاء. فهل تقدم هذه النتيجة دلالة كافية لنقض النظرية؟ علل إجابتك إحصائيا.

٣) نفّذنا عددا من تجارب علم النفس كما يلي: استدرجنا فأرا إلى نهاية حاجز يتفرع منه ممران يقود كل منهما إلى باب. وهدف التجربة أساسا هو تحديد ما إذا كان للفأر قدرة على تفضيل أحد الممرين. في تجربة مؤلفة من 6 محاولات لوحظت النتائج التالية:

المحاولة	1	2	3	4	5	6
الباب الذي اختير	2	1	2	2	2	2

أ - عبر عن الفرضية التي تود اختبارها.
ب - ليكن X عدد المرات التي يختار الفأر فيها الباب الثاني، فما هي قيمة α في هذا الاختبار إذا احتوت منطقة الرفض على $X = 0$ و $X = 6$؟
جـ - ما هي قيمة β من أجل الفرضية البديلة $p = 0.8$؟

٤) سجلنا عدد المآخذ الكهربائية التي تحوي عيبا صناعيا في كل من خطي إنتاج مختلفين A و B، وذلك يوميا ولمدة عشرة أيام فحصلنا على النتائج التالية:

اليوم	1	2	3	4	5	6	7	8	9	10
الخط A	172	165	206	184	174	142	190	169	161	200
الخط B	201	179	159	192	177	170	182	179	169	210

إذا علمت أن حجم الإنتاج الكلي هو نفسه بالنسبة للخطين. قارن عدد القطع المعيبة الناتجة عن الخطين كل يوم، وليكن X عدد الأيام التي يكون فيها B متجاوزا لـ A، فهل تقدم هذه المعلومات دلالة كافية للقول بأن الخط B ينتج في المتوسط قطعا معيبة أكثر من A؟ إعرض الفرضية التي ستختبرها واستخدم X إحصاء لهذا الاختبار.

(٤ ـ ٦) توزيع بواسون *

لتوزيع بـواسون مجالات تطبيق واسعة، فهـو يقدم، على وجه العموم، نمـوذجا جيدا للمعلـومات الاحصـائية التي تأخذ شكل تعـداد لحوادث نادرة الـوقوع. ويمثل المتغير العشوائي البواسوني، X مثلا، عدد «الحوادث النادرة» الملحـوظة في وحدة قياس معينة، زمنا كانت أم مسـافة أم مساحة أم حجما. ونوضح بالأمثلـة التالية، التي نطبق فيها عادة توزيع بواسون:

١ ــ ليكن X عـدد المكـالمات الهاتفية، في شركـة معينـة، كل خمس دقـائق من الفترة الممتدة بين الساعة الثانية عشرة ظهرا والساعة الثانية بعد الظهر.
٢ ـ ليكن X عدد قوالب الزبدة المباعة خلال يوم في محلات بيع المواد الغذائية.
٣ ـ ليكن X عدد الأعطـال الأسبوعية الناشئـة عن العجلات في أسطول من شـاحنات النقل البري.
٤ ـ ليكن X عدد الجسيمات الصادرة في الثانية عن كمية من مادة مشعة.
٥ ـ ليكن X عدد الأخطاء المطبعية في الصفحة عبر صفحات كتاب معين.
٦ ـ ليكن X عدد الالكترونات التي يُصدرها مهبط مسخن في فترة زمنية محددة.
٧ ـ ليكن X عــدد ذرات الغاز في منطقة جـزئية صغــيرة v من وعــاء مليء بهذا الغاز حجمه V.
٨ ـ ليكن X عدد حوادث السيارات في مدينة كبيرة خلال يوم.
٩ ـ ليكن X عدد البكتريا الموجودة في مم^3 من وعاء يحتوي على سائل معين.

وتكفي هذه الأمثلة لتوضيح مدى تنوع واتساع تطبيقات التوزيع البواسوني.

(٤ ـ ٦ ـ ١) دالة الاحتمال لتوزيع بواسون*

يمكن استنتاج دالـة الاحتمال لتوزيع بـواسون كحالـة حدية (أو كنهـاية) لـدالة الاحتمال للتوزيع الثنائي. فلنفرض أن عدد التكرارات n يسعى في اتجاه أن يصبح كبيرا

* للقراءة فقط.

جدا وأن p يسعى في اتجاه أن يصبح صغيرا جـدا ، وبحيث يبقى جـداؤهما $n\,p$ مساويا لعدد ثابت λ ، مثلا ، ولنكتب دالة الاحتمال للتوزيع الثنائي على الشكل :

$$f(x) = \binom{n}{x}\; p^{x}\, q^{n-x} = \frac{n!}{x!\,(n-x)!}\; p^{x}\, \frac{(1-p)^{n}}{(1-p)^{x}}$$

$$= n(n-1)\ldots(n-x+1)\, \frac{\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n}}{x!\left(1-\frac{\lambda}{n}\right)^{x}}$$

$$= \frac{n}{n}\,\frac{n-1}{n}\ldots\frac{n-x+1}{n}\;\frac{\lambda^{x}\left(1-\frac{\lambda}{n}\right)^{n}}{x!\left(1-\frac{\lambda}{n}\right)^{x}}$$

ومن أجل قيمة كبيرة جدا لـ n ، وقيمة لـ x ثابتة وصغيرة بالمقارنة مع n ، تكون كل من النسب $\frac{n-1}{n}$ ، . . . ، $\frac{n-x+1}{n}$ و $\left(1-\frac{\lambda}{n}\right)^{x}$ قـريبة جـدا من الواحـد ، ويمكن كتابة $f(x)$ بصورة تقريبية على الشكل :

$$f(x) \doteq \frac{\lambda^{x}}{x!}\left(1-\frac{\lambda}{n}\right)^{n}$$

حيث $\doteq$ تعني يساوي تقريبا . وإحدى نتائج التحليل الرياضي المعروفة هي أن المقدار $\left(1+\frac{1}{n}\right)^{n}$ يسعـى إلى عـدد ثـابت $e = 2.7183\ldots$ عنـدمـا تسعى n إلى اللانهايـة . وأن $\left(1-\frac{\lambda}{n}\right)^{n}$ يسعى إلى العدد $e^{-\lambda}$ وهكـذا تصبـح الصيغـة الحدية لعبارة $f(x)$ كما يلي :

$$f(x) = e^{-\lambda}\,\frac{\lambda^{x}}{x!}\;;\; x = 0, 1, 2, \ldots$$

وهي صيغة دالة الاحتمال لتوزيـع بواسون . وستعطي هذه الصيغـة احتمالات مساوية تقريبـا لتلك التي تعطيها دالـة الاحتمال للتوزيع الثنائي شريطـة أن يكون n كبيرا جدا ويكون الجداء $n\,p$ صغيرا نسبيا (نطلب عادة أن يكون $n\,p < 5$).

ويبرهن أن كلا من متوسط توزيع بواسون وتباينه يساوي λ. أي أن

$$E(X) = V(X) = \lambda$$

ويمكن اعتبار هذه الخاصة كخاصة مميزة ينفرد بها التوزيع البواسوني من بين التوزيعات المنفصلة جميعها. وإذا وجدنا في مجتمع من القياسات أن متوسطه وتباينه قريبان جدا من بعضهما فإن ذلك يحفزنا على الاعتقاد بأن أفضل نموذج احتمالي مناسب لهذا المجتمع قد يكون النموذج البواسوني.

مثال (٤ ـ ٨)

يتلقى عامل الهاتف في شركة معينة المكالمات الهاتفية بمعدل مكالمتين في الدقيقة.

أ ـ ما احتمال ألا يتلقى العامل أية مكالمة خلال فترة دقيقة؟

ب ـ ما احتمال وصول مكالمتين خلال فترة دقيقة واحدة؟

جـ ـ ما احتمال ألا يتلقى أية مكالمة خلال فترة خمس دقائق؟

الحـــل

لتحديد دالة الاحتمال لتوزيع بواسون تكفي معرفة λ، وهو يمثل متوسط عدد الحوادث التي تقع في وحدة قياس معينة، هي في مثالنا هنا وحدة قياس زمني وتساوي دقيقة واحدة. إذا $\lambda = 2$، على أساس أن وحدة قياس الزمن هي الدقيقة. ولكن كم تصبح λ لو أن وحدة قياس الزمن أصبحت 5 دقائق بدلا من دقيقة واحدة؟ والجواب واضح، لأنه إذا كان متوسط عدد المكالمات يساوي 2 لكل دقيقة فهو يساوي 10 لكل خمس دقائق. وتجدر ملاحظة أن X في توزيع بواسون يمثل عدد الحوادث التي تقع في وحدة قياس. و $f(x) = P(X = x)$ هو احتمال أن تقع x حادثة في وحدة قياس. وهذا يشير إلى ضرورة التعرف على وحدة القياس أو تحديدها ومن ثم حساب λ وكتابة صيغة دالة الاحتمال الموافقة. وبعد ذلك التعبير عن السؤال المطلوب بدلالة x وحساب الاحتمال المطلوب.

وفي مثالنا وحدة القياس هي الدقيقة بالنسبة للسؤالين أ و ب. وتكون λ كما ذكرنا مساوية لـ 2، فنكتب دالة الاحتمال كما يلي:

$$f(x) = e^{-2} \frac{2^x}{x!} \; ; \; x = 0, 1, 2, \ldots$$

أ ـ المطلوب هو $P(X=0)$ أي $f(0)$ وبتعويض x بصفر في الدالة $f(x)$ نجد :

$$f(0) = e^{-2}\,\frac{2^0}{0!} = e^{-2} = 0.135 \,.$$

ب ـ المطلوب هو $P(X=2)$ ، أي $f(2)$. وبتعويض x بـ 2 في $f(x)$ نجد :

$$f(x) = e^{-2}\,\frac{2^2}{2!} = 2\,e^{-2} = 0.270$$

جـ ـ باعتبار وحدة القياس الزمني الآن هي «خمس دقائق»، تصبح $\lambda = 10$ وتصبح دالة الاحتمال كما يلي :

$$f(x) = e^{-10}\,\frac{10^x}{x!} \;;\; x = 0, 1, 2, \ldots$$

والمطلوب هو $f(0) = P(X=0)$. وبتعويض x بصفر في هذه الدالة نجد :

$$f(0) = e^{-10}\,\frac{10^0}{0!} = e^{-10} = 0.000045$$

مثال (٤ ـ ٩)

قام بيتمان بتطبيق توزيع بواسون في مسألة فيزيائية مهمة. فقد استخدم دالة الاحتمال البواسونية في تفسير بيان إحصائي تجريبي كان قد جمعه عالمان عظيمان من رواد الفيزياء الذرية هما رذرفورد وجايجر. فقد قاما بتعداد جسيمات α التي انبعثت عن قرص مطلي بالبولونيوم وذلك خلال فترة زمنية تساوي 7.5 ثانية. وسجلا مشاهداتهما في 2608 فترات زمنية متلاحقة. وكان مجموع عدد الجسيمات الملحوظة يساوي 097 10 جسيما. أي أن متوسط عدد الجسيمات الصادرة هو 3.87 جسيما لكل 7.5 ثانية. (أي لكل وحدة قياس حيث وحدة القياس هنا هي 7.5 ثانية). وقد بين بيتمان أنه إذا كانت نظريتهما الذرية صحيحة وكان λ متوسط عدد الجسيمات الصادرة خلال فترة زمنية محددة، فإن X عدد الجسيمات الصادرة خلال فترة زمنية هو متغير عشوائي يخضع لتوزيع بواسون بوسيط (أو معلمة) يساوي λ ، وهكذا إذا استخدمنا 3.87 كأفضل قيمة تخمينية لـ λ متوفرة لنا، فإن نظريتهما الذرية تتنبأ بأن X متغير عشوائي بواسوني دالة احتماله هي :

$$f(x) = e^{-3.87}\,\frac{(3.87)^x}{x!} \;;\; x = 0, 1, 2, \ldots$$

ويبين الجدول (٤ ـ ٥) النتائج الواقعية والنتائج النظرية، والتوافق الملحوظ القائم بين المشاهدات التجريبية والتنبؤات النظرية .

جدول (٤ ـ ٥) القيم الواقعية والقيم النظرية لتجربة رذرفوردوجايجر

	عدد الفترات الزمنية (7.5 ثا) التي صدر خلالها n جسيم α	
n	القيمة الملحوظة	القيمة النظرية (مع تدوير الرقم العشري)
0	57	54
1	203	210
2	383	407
3	525	525
4	532	508
5	408	394
6	273	254
7	139	140
8	45	68
9	27	29
10	10	11
11	4	4
12	0	1
≥13	2	1

مثال (٤ ـ ١٠)

تقع حوادث اصطدام الطرق في منطقة معينة بمعدل حادث واحد لكل يومين .

ا ـ احسب الاحتمالات الموافقة لـ 0، 1، 2، 3، 4، 5، 6 حوادث اصطدام في الأسبوع في تلك المنطقة .

ب ـ ما عدد الاصطدامات الأسبوعية الأكثر احتمالا؟

جـ ـ كم يوما في الأسبوع تتوقع أن يمر بدون اصطدامات؟

الحـــل

ا ـ متوسط عدد الحوادث لكل أسبوع هو $\lambda = 7\ (0.5) = 3.5$. ودالة الاحتمال هي :

$$f(x) = e^{-3.5}\frac{(3.5)^x}{x!} \quad ; x = 0, 1, 2, \ldots$$

ومنه :

$$f(0) = e^{-3.5}\frac{(3.5)^0}{0!} = 0.030 \quad ; \quad f(1) = e^{-3.5}\frac{3.5}{1!} = 0.106$$

$$f(2) = e^{-3.5}\frac{(3.5)^2}{2!} = 0.185 \quad ; \quad f(3) = e^{-3.5}\frac{(3.5)^3}{3!} = 0.216$$

$$f(4) = e^{-3.5}\frac{(3.5)^4}{4!} = 0.189 \quad ; \quad f(5) = e^{-3.5}\frac{(3.5)^5}{5!} = 0.132$$

$$f(6) = e^{-3.5}\frac{(3.5)^6}{6!} = 0.077 .$$

ب ـ من السؤال أ نلاحظ أن عدد الحوادث الأسبوعية الأكثر احتمالا هو $x = 3$.

جـ ـ باعتبار اليوم هو وحدة القياس بدلا من الأسبوع نجد أن $\lambda = 0.5$ وتكون دالة الاحتمال

$$f(x) = e^{-0.5}\frac{(0.5)^x}{x!} \quad ; x = 0, 1, \ldots$$

حيث x الآن هو عدد الحوادث اليومية . ومرور يوم بدون حوادث يعني أن $x = 0$.ولحساب $P(X = 0)$ نعوض x بصفر في دالة الاحتمال فنجد :

$$f(0) = e^{-0.5} = 0.607$$

ونحن الآن أمام مسألة توزيع ثنائي . لنعّرف النجاح بأنه مرور يوم بدون حوادث ، باحتمال النجاح $p = 0.607$، في كل يوم من أيام الأسبوع السبعة . وبأخذ $n = 7$ ، يكون عدد النجاحات هو عدد أيام الأسبوع التي تمر بدون حوادث . إذا رمزنا لهذا العدد بـy، فإن y متغير عشوائي يخضع للتوزيع الثنائي حيث $n = 7$ و $p = 0.607$. والمطلوب هو $E(y)$. وكما نعلم فإن :

$$E(y) = n\,p = 7 \times 0.607 = 4.25$$

وبصورة تقريبية نقول إنه لو أحصينا عدد أيام الأسبوع التي تمر بدون حوادث في تلك المنطقة وذلك لعدد هائل من الأسابيع ، ثم حسبنا متوسط الأعداد التي حصلنا عليها لكان الناتج 4.25 يوما .

تمارين (٤ ـ ٤)

١) تتلقى تحويلة للهاتف المكالمات بين الساعة العاشرة صباحا والثانية عشرة ظهرا بمعدل مكالمتين في الدقيقة. ما هو احتمال ألا تتلقى التحويلة أية مكالمة خلال دقيقة؟ أن تتلقى مكالمتين خلال دقيقة؟ أن تتلقى مكالمتين خلال خمس دقائق؟ ألا تتلقى أية مكالمة خلال خمس دقائق؟

٢) لنفرض أن مساحة صغيرة من زجاجة مجهرية لفحص الدم تحتوي في حالة شخص طبيعي على عشر كريات حمر في المتوسط. ما احتمال أن تتضمن زجاجة من دم شخص طبيعي، في تلك المساحة الصغيرة، أقل من 6 كريات حمر؟ ألا تحوي أية كرية حمراء؟

٣) تتضمن صحيفة يومية في المتوسط ثلاثة أخطاء مطبعية للصفحة الواحدة. ما احتمال أن:

ا ـ تكون الصفحة الأولى خالية من الأخطاء المطبعية؟

ب ـ يوجد ستة أخطاء مطبعية في الصفحة الأخيرة؟

جـ ـ يوجد أكثر من ثلاثة أخطاء مطبعية في صفحة الرياضة.

٤) يبيع مخزن نوعا معينا من الأجهزة الكهربائية بمعدل أربعة في الأسبوع. بافتراض أن عدد الأجهزة المباعة أسبوعيا متغير بواسوني، أوجد عدد الأجهزة التي يجب توافرها في مستودع المخزن في بداية أسبوع بحيث يطمئن صاحب المخزن باحتمال 95% إلى أنه سيلبي جميع الطلبات من هذا النوع من الأجهزة خلال ذلك الأسبوع.

٥) تصل السيارات إلى مرآب في وسط المدينة بمعدل سيارة كل دقيقة. وسيسبب وصول أكثر من أربع سيارات في أي دقيقة أزمة في حركة المرور. كم أزمة تتوقع، في المتوسط، في ساعات العمل الـ 12 في اليوم؟

٦) من بين 150 مباراة في كرة القدم جرت يوم الخميس لم تُسجل أية أهداف في 12 منها. مفترضا توزيع بواسون، كم تعتقد أن يكون متوسط عدد الأهداف للمباراة الواحدة؟

احسب احتمالات أن:

ا - يسجل أقل من هدفين في مباراة معينة.

ب - يسجل أكثر من هدفين ولكن أقل من 5 أهداف في مباراة معينة.

٧) تمر المركبات من نقطة معينة على طريق مزدحم بمعدل 300 مركبة في الساعة. أوجد احتمال ألا تمر أي مركبة خلال دقيقة معينة. ما العدد المتوقع للمركبات التي تمر خلال دقيقتين. أوجد احتمال أن يمر بالفعل هذا العدد المتوقع خلال أي فترة طولها دقيقتان؟

٨) سجل عدد حوادث الاصطدام في منطقة معينة يوميا ولفترة امتدت 1500 يوم، وكانت النتائج كما يلي:

عدد الاصطدامات في اليوم	0	1	2	3	4	5
التكــرار	342	483	388	176	111	0

ما التوزيع النظري الذي يمكن استخدامه نموذجا مناسبا لهذا البيان؟ احسب التكرارات المتوقعة مستخدما التوزيع الاحتمالي النظري بمتوسط يساوي متوسط البيان الإحصائي أعلاه.

٩) في مسح كبير تناول أكثر من 100 000 ولادة تبين أن معدل الاصابة بمرض في العمود الفقري هي 4.12 لكل ألف ولادة. ما احتمال أن نلاحظ في عينة عشوائية من خمسين ولادة:

ا - عدم وجود إصابات؟

ب ـ إصابة واحدة؟
جـ ـ إصابيتين
د ـ أكثر من إصابتين؟
(لاحظ من طريقة اشتقاق التوزيع البواسوني أنه عندما تكون n غير صغيرة، هنا $n = 50$، ويكون احتمال النجاح، صغيرا جدا، هنا $p = .00412$، فيمكن اعتبار التوزيع البواسوني بمتوسط $\lambda = n\,p$ تقريبا جيدا للتوزيع الثنائي.)

(٤ ـ ٧) العينة العشوائية

قلنا إن هدف الإحصاء كعلم هو القيام باستقراء حول خصائص مجتمع اعتمادا على المعلومات التي تحويها عينة مأخوذة من هذا المجتمع. وكل مسألة إحصائية تبدأ بعينة من القياسات أو المشاهدات. وعلى سبيل المثال، عند اتخاذ قرار برفض أو قبول شحنة بضاعة واردة إلى مصنع، وكذلك عند اختبار فرضية تتعلق بفعالية لقاح جديد ضد الزكام، اعتمدنا، في كل حالة، على عينة مأخوذة من مجتمع، ووصفنا العينة بأنها عشوائية، فماذا نبتغي من وصف العينة الاحصائية بأنها عينة عشوائية؟ وفي المقام الأول، متى نقول إن العينة عشوائية؟

ولقد أوضحنا، في مسألة اختبار فرضية، أنه لابد من حساب احتمال الحصول على عينة كالعينة التي بين أيدينا. (العينة التي تمخضت عنها التجربة) فإذا وجدنا أنها من النوع غير المحتمل (احتمال الحصول عليها تحت الفرضية الابتدائية هو احتمال زهيد) نستنتج أن الفرضية الابتدائية غير مبررة، ونرفضها. وإذا وجدنا أن العينة محتملة تماما قلنا إن الدلالات المتوفرة من العينة لا تسمح لنا برفض الفرضية ولذلك نقبلها. والنقطة المهمة التي نريد إبرازها هي أنه لابد لنا من حساب احتمال الحصول على عينة كتلك التي لاحظناها كي نصل إلى استقراء إحصائي، أو نتخذ قرارا إحصائيا. ولا يخفى أن لطريقة أخذ العينة أثرا حاسما في حساب احتمالها. ومن هنا تأتي أهمية كون العينة عشوائية. فالعشوائية هي الخاصية التي ستجعل حساب مثل ذلك الاحتمال ممكنا، لا بل ستجعله سهلا وميسورا. ولكن متى نقول إن العينة عشوائية؟

لنفرض الآن أننا سحبنا عينة تتضمن n قياسا من مجتمع يحوي N قياسا، فما هو عدد العينات المختلفة التي يمكن الحصول عليها؟ إن هذا العدد، كما نعلم، هو عدد متوافقات N شيئا مأخوذ n منها في وقت واحد، أي

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

وفي الفصل الثالث علقنا على مسألة الاختيار العشوائي لعنصر من مجموعة تتضمن N عنصرا، فقلنا إن عشوائية الاختيار تعني أن لكل من العناصر الـ N، الفرصة نفسها في أن يكون العنصر الذي يقع عليه الاختيار. أي أن احتمال الاختيار هو $1/N$ لكل عنصر من العناصر الـ N في المجموعة التي نختار منها. وهذا تعبير كمي عن طريقة اختيار نطمئن فيها إلى عدم إمكانية وجود أي شكل من أشكال التحيز لعنصر دون آخر. وسنطبق الفكرة نفسها لتعريف عشوائية العينة.

تعريف العينة العشوائية

عند سحب عينة حجمها n من مجتمع منته يتضمن N عنصرا، نقول إن العينة عشوائية، إذا كان لكل من العينات الـ $\binom{N}{n}$ الممكنة الفرصة نفسها في أن تكون العينة الملحوظة. أي إذا كان احتمال الحصول على أي منها هو $\dfrac{1}{\binom{N}{n}}$

وفي معظم التطبيقات الاحصائية يكون المجتمع لانهائيا (أي أن عدد عناصره غير محدود)، وتجريدا ذهنيا أكثر منه عناصر محسوسة. لنأخذ، مثلا، حالة القيام بقياس ثابت فيزيائي في تجربة مخبرية، ولنفرض أننا كررنا التجربة نفسها عشر مرات، فعندئذ ننظر إلى القياسات العشرة الناتجة على أنها عينة من مجتمع افتراضي هو ذلك المجتمع من القياسات التي كنا سنحصل عليها لو أننا قمنا، وبصورة مستقلة، بتكرار التجربة نفسها مرة بعد أخرى إلى ما لانهاية له. وكل قياس من القياسات العشرة هو في الواقع متغير عشوائي قائم بذاته، ويوافقه بالطبع مجتمع من القياسات. وعشوائية العينة تضمن لنا أن المجتمعات العشرة من القياسات، هي، في الحقيقة، مجتمع واحد، وفوق ذلك تضمن لنا أن هذه المتغيرات العشرة تتحول، أو تعمل، مستقلة بعضها عن بعض. وبعبارة مبسطة نقول إنه كي نحصل على عينة عشوائية من n قياسا، ما علينا

إلا أن نكرر التجربة، تحت نفس الشروط والظروف، n مرة. وبطريقة تسمح لنا بالقول إن هذه التكرارات الـn مستقلة فيما بينها، أي لا يمكن أن يكون لنتيجة أي تكرار منها أثر سلبي أو إيجابي على ما يمكن أن تكونه نتيجة تكرار آخر.

ولو أمعنا النظر فيما نقوله وتذكرنا، على سبيل المقارنة، ما قلناه عند تعريف تجربة ثنائية، لوجدنا أن التكرارات الـn لتجربة ثنائية إنما تمثل تماما عينة عشوائية حجمها n من مجتمع القياسات الموافق لمتغير ثنائي نقطي (متغير بيرنويللي). فثبات قيمة p من تكرار إلى آخر يلخص شرط ثبات ظروف التجربة، وأننا نكرر التجربة نفسها مرة بعد أخرى، وهو الشرط الأول من شرطي عشوائية العينة، أما استقلال التكرارات بعضها عن بعض فيستكمل الشرط الثاني المطلوب. وسنستخدم هذه الملاحظة المهمة للوصول إلى طريقة حسابات تقريبية، إلا أنها سهلة ومفيدة، في الفصل القادم. وتسمى «طريقة تقريب التوزيع الثنائي بالتوزيع الطبيعي».

(٤ ـ ٨) المعاينة بدون إرجاع والتوزيع فوق الهندسي

عند سحب عينة عشوائية من مجتمع منته (يتضمن عددا محدودا من العناصر)، إذا سحبنا العنصر وسجلنا نتيجة السحب ثم أعدنا العنصر إلى المجتمع قبل سحب عنصر آخر، سُميت المعاينة «معاينة مع الإعادة». أما إذا احتفظنا بالعنصر المسحوب وقمنا بالسحب التالي من العناصر الباقية في المجتمع، أي لم نقم بإعادة العنصر المسحوب إلى المجتمع، سُميت المعاينة «معاينة بدون إعادة». وإذا تضمن المجتمع N عنصرا، مثلا، فسيبقى المجتمع على حاله، بدون تغيير، في الحالة الأولى، إذ يجري، على الدوام، سحب عنصر من بين N عنصرا، هي مجمل عناصر المجتمع. ومن الواضح أن نتيجة كل سحب ستكون مستقلة عن نتيجة أي سحب آخر. وعند سحب عينة حجمها n تكون عمليات السحب الـn تكرارات مستقلة للتجربة نفسها، مما يتفق تماما مع شروط التوزيع الثنائي. أما إذا كانت المعاينة بدون إعادة فإن نتيجة كل سحب ستتأثر بنتائج جميع عمليات السحب السابقة. ولا تشكل عمليات السحب الـn تكرارات مستقلة بعضها عن بعض، ولا ينطبق عليها بالتالي التوزيع الثنائي. وإذا

كان الأثر زهيدا، n صغيرة جدا بالنسبة لـN، أي إذا كان الحيدان عن شروط التوزيع الثنائي في حدود طفيفة، فيمكن تطبيق التوزيع الثنائي كتقريب جيد. وفيما عدا ذلك لابد لنا من التفكير في توزيع يتلاءم وشروط المعاينة. وسنجد أن التوزيع فوق الهندسي هو التوزيع الملائم لمعاينة بدون إعادة فما هو التوزيع فوق الهندسي، وما هي الحالات التي نلجأ فيها إلى هذا التوزيع؟

لنفرض أن مجتمعا يتضمن N عنصرا من بينها N_1 عنصرا يتصف بصفة معينة، A مثلا، والعناصر الباقية وعددها $N - N_1$ لا تتصف بالصفة A. إذا سحبنا من هذا المجتمع، وبدون إعادة، عينة عشوائية حجمها n، وعرفنا المتغير العشوائي X بأنه عدد عناصر العينة التي تتصف بالصفة A. فالتوزيع الاحتمالي لـX يسمى التوزيع فوق الهندسي. وللوصول إلى صيغة هذا التوزيع $f(x)$، علينا أولا تحديد القيم الممكنة لـX، ثم الاجابة على السؤال التالي:

ما احتمال أن يكون X مساويا لقيمة ما x، حيث ترمز x لأي قيمة من القيم الممكنة. أي ما هو احتمال الحادثة $X = x$، ونكتب:

$$f(x) = P(X = x)$$

ومن الواضح أن القيم الممكنة لـX هي:

$$0,1, 2,\ldots,\min(N_1, n)$$

حيث يعني الرمز $\min(N_1, n)$ أصغر العددين N_1, n. فالقيمة x لا يمكن أن تتجاوز n باعتبارها تمثل جزءا من العينة، ولا يمكنها، على الوجه الآخر، أن تتجاوز N_1 لأن العينة، وهي جزء من المجتمع، لا يمكن أن تتضمن من عناصر الصفة A أكثر مما في المجتمع من هذه العناصر.

ولحساب $f(x)$، نتذكر أن التجربة هي سحب عينة عشوائية بدون إعادة حجمها n من المجتمع الموصوف أعلاه، ثم تسجيل x عدد العناصر في هذه العينة التي تتصف بالصفة A. وفضاء العينة هو مجموعة النتائج الممكنة، أي مجموعة كل العينات الممكنة وعددها $\binom{N}{n}$. وبما أن العينة عشوائية فلكل منها الاحتمال نفسه وهو $1 / \binom{N}{n}$. أي

أننا هنا في حالة نموذج الاحتمالات المتساوية. ويكون احتمال الحادثة $X = x$ هو عدد الحالات الملائمة، . أي عدد العينات التي تتضمن x من عناصر الصفة A ، مقسوما على عدد الحالات الممكنة وهو $\binom{N}{n}$. ولكن عدد الحالات الملائمة هو عدد إمكانات اختيار x من N_1 و $n - x$ من $N - N_1$ ، أي $\binom{N_1}{x}\binom{N - N_1}{n - x}$ وهكذا نجد :

$$f(x) = \frac{\binom{N_1}{x}\binom{N - N_1}{n - x}}{\binom{N}{n}} \; , \; x = 0, 1, 2, \ldots , \min \; (N_1 , n) .$$

وهي صيغة التوزيع فوق الهندسي .

ويمكن البرهان على أن متوسط هذا التوزيع :

$$E(X) = n\frac{N_1}{N} .$$

أي n مضروبا بنسبة تواجد عناصر الصفة A في المجتمع . كما يمكن ا لبرهان على أن تباين التوزيع $V(X)$ هو:

$$V(X) = \frac{N - n}{N - 1} \; n\frac{N_1}{N}\left(1 - \frac{N_1}{N}\right)$$

مثال (٤ ـ ١١)

خضع اثنا عشر مريضا بالتهاب القصبات إلى تجربة طبية فاختير ستة منهم بصورة عشوائية وأعطوا المعالجة A بينما أعطي الباقون المعالجة B. بكم طريقة يمكن توزيع الاثني عشر مريضا على المعالجتين؟ وإذا كان أربعة من المرضى يعانون من ارتفاع ضغط الدم والآخرون طبيعيون بالنسبة لمعدل ضغط الدم ، فما احتمالات :

أ ـ أن يخضع ذوو الضغط المرتفع جميعهم للمعالجة A؟

ب ـ أن يخضع ذوو الضغط المرتفع جميعهم للمعالجة نفسها؟

جـ ـ أن يوجد واحد على الأقل من ذوي الضغط المرتفع في كل معالجة؟

الحـــل

عدد إمكانات توزيع المرضى على المعالجتين هو $\binom{12}{6}$ لأن اختيار 6 وتخصيصهم للمعالجة A يعني بطبيعة الحال تخصيص الستة الباقين للمعالجة B.

$$\binom{12}{6} = \frac{12!}{6!\ 6!} = \frac{12.\ 11.\ 10.\ 9.\ 8.\ 7}{6.\ 5.\ 4.\ 3.\ 2.\ 1} = 924$$

أ ـ لدينا هنا $N = 12$، $N_1 = 4$، $N - N_1 = 8$، $n = 6$، وليكن X عدد المصابين بالضغط المرتفع الذين يجري تخصيصهم للمعالجة A. فالتوزيع الاحتمالي لـ X هو التوزيع فوق الهندسي.

$$f(x) = \frac{\binom{4}{x}\binom{8}{6-x}}{\binom{12}{6}} \ ; \ x = 0, 1, 2, 3, 4\ .$$

والمطلوب في أ هو احتمال $x = 4$، و هو $f(4)$. وهكذا نكتب :

$$P(X=4) = f(4) = \frac{\binom{4}{4}\binom{8}{2}}{\binom{12}{6}} = \frac{1}{33}$$

ب ـ السؤال هنا يعني أن يكون $x = 4$، جميع ذوي الضغط المرتفع خاضعون للمعالجة A أو $x = 0$، جميع ذوي الضغط المرتفع خاضعون للمعالجة B. وهكذا نكتب :

$$P(X = 4 \text{ أو } 0) = P(X=0) + P(X=4) = f(0) + f(4)$$
$$= \frac{\binom{4}{0}\binom{8}{6}}{\binom{12}{6}} + f(4) = \frac{2}{32}$$

جـ ـ هذه الحادثة هي الحادثة المتممة للحادثة المذكورة في ب واحتمالها يساوي

$$1 - \frac{2}{32} = \frac{31}{33}\ .$$

أو إذا حسبنا المطلوب بصورة مباشرة نجده :

$$P(3 \text{ أو } 2 \text{ أو } 1) = f(1) + f(2) + f(3) = \frac{31}{33}$$

مثال (٤ ـ ١٢)

عجنة تتضمن 3 عناصر من النوع أ و 5 عناصر من النوع ب . سحبنا عينة عشوائية* من أربعة عناصر من هذه العجنة . ليكن X عدد العناصر في العينة من النوع أ .

١ ـ اكتب صيغة التوزيع الاحتمالي لـ X ، وجدولا يتضمن كل قيمة من القيم الممكنة لـ X ، والاحتمال المقابل لها .

٢ ـ احسب $E(X)$ و $V(X)$ باستخدام الجدول (أي بتطبيق التعريف مباشرة) . ثم باستخدام الصيغتين المعطاتين لمتوسط التوزيع وتباينه وقارن النتيجتين .

٣ ـ إذا كان عمر كل عنصر من النوع أ سنتين ، وعمر كل عنصر من النوع ب خمس سنوات ، فما هي القيمة المتوقعة لعمر العينة؟ وما هو تباين عمر العينة؟

الحل

١ ـ يتضمن المجتمع ثمانية عناصر، أي $N = 8$ منها $N_1 = 3$ عناصر من النوع أ . إذا سحبنا عينة عشوائية من أربعة عناصر، $n = 4$ ، فيكون التوزيع الاحتمالي لـ X هو، بوضوح ، التوزيع فوق الهندسي :

$$f(x) = \frac{\binom{3}{x}\binom{5}{4-x}}{\binom{8}{4}} , \quad x = 0, 1, 2, 3 .$$

والجدول المطلوب هو:

x	0	1	2	3
$f(x)$	$\frac{1}{14}$	$\frac{6}{14}$	$\frac{6}{14}$	$\frac{1}{14}$

* عبارة «سحبنا عينة عشوائية» ستعني دائما معاينة بدون إعادة ما لم يُذكر غير ذلك .

٢ ـ باستخدام الجدول :

$$E(X) = 0 \times \frac{1}{14} + 1 \times \frac{6}{14} + 2 \times \frac{6}{14} + 3 \times \frac{1}{14} = \frac{3}{2}$$

$$E(X^2) = 0^2 \times \frac{1}{14} + 1^2 \times \frac{6}{14} + 2^2 \times \frac{6}{14} + 3^2 \times \frac{1}{14} = \frac{39}{14}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{39}{14} - \frac{9}{4} = \frac{156 - 126}{56} = \frac{15}{28} .$$

وعلى الوجه الآخر، كان يمكن التعويض في صيغتي متوسط التوزيع وتباينه لنجد :

$$E(X) = n\frac{N_1}{N} = 4 \times \frac{3}{8} = \frac{12}{8} = \frac{3}{2} ,$$

$$V(X) = \frac{N - n}{N - 1} \, n\frac{N_1}{N}\left(1 - \frac{N_1}{N}\right)$$

$$= \frac{8-4}{8-1} \times 4 \times \frac{3}{8}\left(1 - \frac{3}{8}\right)$$

$$= \frac{4}{7} \times \frac{12}{8} \times \frac{5}{8} = \frac{15}{28}$$

وهو بالضبط ما وجدناه بالاستخدام المباشر للتعاريف .

٣ ـ لنرمز لعمر العينة بـ Y فيكون :

$$Y = 2X + 5(4 - X) = 20 - 3X$$

والمطلـوب $E(Y)$ و $V(Y)$. ولكن نعلم من خواص التـوقع وخواص التباين أن :

$$E(Y) = E(20 - 3X) = 20 - 3E(X)$$

$$= 20 - 3 \times \frac{3}{2} = \frac{31}{2} .$$

$$V(Y) = V(20 - 3X) = 9V(X) = 9 \times \frac{15}{28} = \frac{135}{28} .$$

(٤ ـ ٩) توزيع $\bar{x}$ متوسط عينة من مجتمع منته

نحتاج في عمليات الاستقراء الإحصائي، أشد ما نحتاج، إلى التوزيعات الاحتمالية لخصائص العينة. ما كان منها مقياسا للنزعة المركزية، أو ما كان منها مقياسا للتشتت. وفي الطليعة منها جميعا نجد متوسط العينة $\bar{X}$.

لنفترض أن لدينا مجتمعا منتهيا فيه N عنصرا. إذا سحبنا من هذا المجتمع عينة عشوائية حجمها n ، ورمزنا لمتوسطها بـ $\bar{X}$ ، فماذا نقصد بتوزيع $\bar{x}$ ؟ التوزيع الاحتمالي هو، عمليا، وصف وتحديد لبنية مجتمع القياسات الموافق لمتغير عشوائي، والمتغير العشوائي، كما نعلم، ينبغي أن يكون معرفا على فضاء عينة، وفضاء العينة هو مجموعة النتائج الممكنة لتجربة، فأين التجربة هنا وما فضاء العينة؟ من الواضح أن التجربة هي سحب عينة عشوائية حجمها n من مجتمع يتضمن N عنصرا، وبالتالي ليس فضاء العينة هنا إلا مجموعة كل العينات التي يمكن الحصول عليها. ويجدر الانتباه إذا إلى أن ما يؤخذ في الاعتبار ليس عينة واحدة، سحبناها وحسبنا متوسطها، ولكن مجمل العينات التي كان يمكن أن نحصل عليها لو أننا كررنا تجربة السحب مرة بعد أخرى. وهنا نضع اليد من جديد على الطبيعة التكرارية للمسألة الاحتمالية والمسألة الاحصائية. صحيح أننا نعتمد على المعلومات التي تقدمها العينة للقيام باستقراء حول المجتمع الذي جاءت منه. ولكننا لا نعتمد على هذه المعلومات كقطعة معزولة قائمة بذاتها، وإنما نعتمد عليها، في سياق شريط متكامل، أو جزء من صورة متكاملة، تتضمن العينة التي بين أيدينا وغيرها من العينات التي كنا سنحصل عليها لو أننا كررنا أخذ عينة ثانية وثالثة وهلمجرا. والتوزيع الاحتمالي للعينة أو، على وجه التحديد، لخاصة من خصائصها، ولنقُل $\bar{X}$ مثلا، هو الذي يقدم وصفا لمحتويات تلك الصورة المتكاملة. فمثلا، ما احتمال أن تكون قيمة $\bar{X}$ أكبر من عدد معين؟ أو بعبارة عملية، ما نسبة العينات التي يزيد متوسطها على عدد معين؟ وذلك من بين كل العينات الممكنة؟ ومن خلال التوزيع الاحتمالي لـ $\bar{X}$ يمكننا الإجابة على أسئلة كهذه، كما يمكننا، بصورة عامة، الحكم بأن العينة التي حصلنا عليها هي، مثلا من النوع غير

المحتمل، أو من النوع المحتمل، الأمر الذي ساعدنا عند مناقشة مسألة اختبار فرضية على اتخاذ موقف إحصائي من الفرضية، وكان له الدور الأساسي في بلورة مثل ذلك الموقف.

وللوصول إلى التوزيع الاحتمالي للمتغير العشوائي $\bar{X}$ نحسب قيمته عند كل نقطة عينة أي لكل عينة من العينات الـ$\binom{N}{n}$ الممكنة. ويكون الاحتمال الموافق لكل من القيم المختلفة لـ $\bar{X}$ هو $1/\binom{N}{n}$ مضروبا بعدد المرات الذي تكرر فيه ظهور تلك القيمة. وسنوضح الطريقة بمثال.

مثال (٤ ـ ١٣)

لدينا المجتمع من الأرقام 1,2,3,4,5,6. اكتب التوزيع الاحتمالي لـ $\bar{X}$ متوسط عينة حجمها 2 نسحبها عشوائيا من هذا المجتمع. وارسم المدرج الاحتمالي لهذا التوزيع.

الحل

عدد نقاط العينة، أي عدد كل العينات الممكنة هو $\binom{6}{2} = 15$ والاحتمال الموافق لكل منها هو 1/15، وذلك وفقا لتعريف العينة العشوائية. والجدول (٤ ـ ٦) يبين العينات المختلفة الممكنة والاحتمال الموافق لكل منها، والقيمة التي يأخذها المتغير العشوائي $\bar{X}$ في كل نقطة عينة، أي قيمة المتوسط الحسابي للعينة.

ومن هذا الجدول يمكننا بسهولة وضع جدول التوزيع الاحتمالي المطلوب، وهو يتضمن القيم المختلفة لـ $\bar{X}$ والاحتمال الموافق لكل منها. ونلاحظ أن القيم المختلفة لـ $\bar{X}$ هي

$$1.5,\ 2,\ 2.5,\ 3,\ 3.5,\ 4,\ 4.5,\ 5,\ 5.5$$

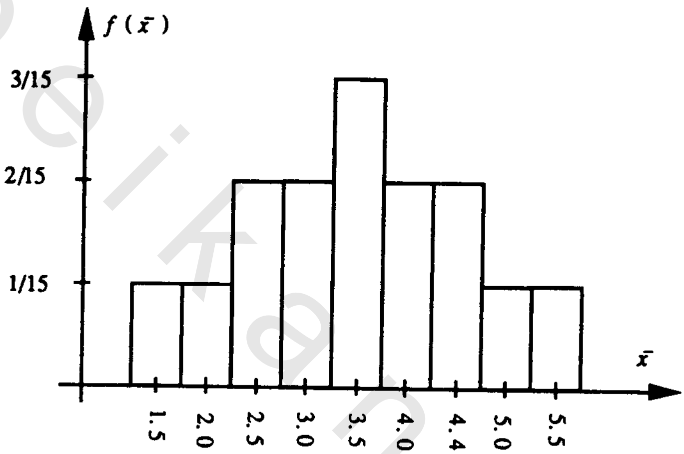
والاحتمال الموافق لـ 2.5 مثلا، هو 1/15 مضروبا بعدد المرات التي تكرر فيها ظهور 2.5 كمتوسط أي:

$$P(\bar{X} = 2.5) = f(2.5) = \frac{1}{15} \times 2 = \frac{2}{15}$$

(انظر الجدول ٤ ـ ٧).

جدول (٤ ـ ٧) التوزيع الاحتمالي لـ $\bar{X}$

$\bar{x}$	1.5	2	2.5	3	3.5	4	4.5	5	5.5
$f(\bar{x})$	1/15	1/15	2/15	2/15	3/15	2/15	2/15	1/15	1/15



شكل (٤ ـ ٥): المدرج الاحتمالي لتوزيع $\bar{x}$ في المثال (٤ ـ ١٣)

جدول (٤ ـ ٦)

قيمة $\bar{X}$	الاحتمال الموافق	العينات الممكنة (نقاط العينة)
1.5	1/15	1,2
2	1/15	1,3
2.5	1/15	1,4
3	1/15	1,5
3.5	1/15	1,6
2.5	1/5	2,3
3	1/15	2,4
3.5	1/15	2,5
4	1/15	2,6
3.5	1/15	3,4
4	1/15	3,5
4.5	1/15	3,6
4.5	1/15	4,5
5	1/15	4,6
5.5	1/15	5,6

لاحظ أن المدرج الاحتمالي لتوزيع $\bar{X}$ متنـاظر تماما في هذا المثال. وهـو مقبب في الوسط ويتناقص تدريجيا على اليمين وعلى اليسار.

(٤ـ٩ـ١) خواص، $\bar{X}$ ، متوسط عينة عشوائية حجمها n مأخوذة من مجتمع حجمه N

لنرمز بـ μ (ميو) لمتوسط المجتمع، وبـ σ^2 (سيجما مربع) لتباين المجتمع. أي:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \;,\; \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{N}\left[\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{1}^{N} x_i\right)^2}{N}\right]$$

فيمكن البرهان على الخواص التالية:

١ ـ القيمة المتوقعة لـ $\bar{X}$ تساوي تماما متوسط المجتمع أي:

$$E(\bar{X}) = \mu$$

٢ ـ تباين $\bar{X}$ ولنرمز له بـ $\sigma^2_{\bar{X}}$ معطى بالعلاقة التالية:

$$V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{N-n}{N-1}\,\frac{\sigma^2}{n}$$

لاحظ أن تباين $\bar{X}$ الذي يعبر عن تغير قيمة $\bar{X}$ من عينة لأخرى أصغر بكثير من تبـاين المجتمـع. ففي الطـرف الأيمن من عبـارة $V(\bar{X})$ نجـد تبـاين المجتمع σ^2 مقسوما على حجم العينة n، وفوق ذلك أخذنـا جزءا من $\frac{\sigma^2}{n}$ لأن $\frac{N-n}{N-1}$ أصغـر من الواحد. ويكفي أن ننظر إلى المثـال (٤ ـ ١٣) السابق لنجد أن قيم المجتمع تختلف عن بعضها بمقدار الـواحد الصحيح ولكن متوسطـات عينات حجمها 2 مأخوذة من هذا المجتمع لا تختلف عن بعضها إلا بمقدار النصف.

وإذا كـــان حجم العينـة n صغيرا جـدا بـالنسبـة إلى حجم المجتمع N تصبح النسبـة $\frac{N-n}{N-1}$ ، وتسمى عامـل التصحيح في مجتمع منته، قـريبـة جـدا من الواحـد ويصبح تباين متوسط العينة مساويا تقـريبا لتباين المجتمع مقسوما على حجم العينة n. وتتضح هنا نقطتان مهمتان:

(١) يمكن التحكم بتباين $\bar{X}$ وجعله صغيرا من خلال زيادة حجم العينة n. أي أن حجم العينة n يشكل صمام أمان تجدر الاستفادة منه للوصول إلى قرارات وتنبؤات إحصائية جيدة. إلا أن مقدرتنا على استخدام صمام الأمان هذا منوطة بالاستعداد لبذل كافة الجهود والنفقات التي يتطلبها أخذ عينة أو يتطلبها القيام بتجربة إحصائية. والمسألة هنا تصبح مسألة اقتصادية إذ نريد، لقاء نفقة معينة، الوصول إلى قرارات أو تنبؤات سليمة، حول خصائص مجتمع، استنادا إلى عينة نأخذها من هذا المجتمع. وتهدف نظرية الاحصاء إلى تقديم طرق كفؤة، تسمح لنا القيام باستقراءات جيدة، من خلال عينات صغيرة نسبيا.

(٢) كلما كان المجتمع المدروس أقل تجانسا (تباينه σ^2 كبير) اضطررنا إلى زيادة حجم العينة حتى نحافظ على حد مُرض لسلامة وجودة القرار أو التنبؤ الاحصائي.

مثال (٤ ـ ١٤)

في المثال (٤ ـ ١٣) احسب متوسط المجتمع μ وتباينه σ^2. ثم احسب $E(\bar{X})$ و $V(\bar{X})$ وتحقق من صحة العلاقات المذكورة كخواص للمتوسط.

الحل

متوسط المجتمع:

$$\mu = \frac{\sum_{1}^{N} x_i}{N} = \frac{\sum_{1}^{6} x_i}{6} = \frac{1+2+3+4+5+6}{6} = 3.5$$

x_i^2	1	4	9	16	25	36	91
x_i	1	2	3	4	5	6	21

تباين المجتمع:

$$\sigma^2 = \frac{1}{N}\left[\sum_{1}^{N} X_i^2 - \frac{\left(\sum_{1}^{N} X_i\right)^2}{N}\right]$$
$$= \frac{1}{6}\left[91 - \frac{(21)^2}{6}\right] = 2.917$$

$$E(\bar{X}) = \sum_{\bar{x}} \bar{x} \cdot f(\bar{x}) = 1.5 \times \frac{1}{15} + 2 \times \frac{1}{15} + \ldots + 5.5 \times \frac{1}{15} = 3.5 = \mu$$

$$V(\bar{X}) = E(\bar{X})^2 - [E(\bar{X})]^2$$
$$E(\bar{X}^2) = \sum_{\bar{x}} \bar{x}^2 f(\bar{x}) = 1.5^2 \times \frac{1}{15} + 2^2 \times \frac{1}{15} + \ldots + 5.5^2 \times \frac{1}{15} = 13.417$$
$$V(\bar{X}) = 13.417 - 12.25 = 1.167$$

ومن جهة أخرى نجد بتطبيق العلاقة :

$$V(\bar{X}) = \frac{N-n}{N-1}\frac{\sigma^2}{n} = \frac{6-2}{6-1}\frac{2.917}{2} = 1.167$$

وهي النتيجة نفسها التي وجدناها باستخدام التوزيع $f(\bar{x})$ وتعريف التباين .

تمارين (٤ ـ ٥)

١) حظيرة فيها 11 حيوانا منها سبع إناث وأربعة ذكور. اخترنا أربعة حيوانات من الحظيرة عشوائيا . ما احتمالات أن تتضمن :

ا ـ ثلاثة ذكور؟

ب ـ حيوانا واحدا على الأقل من كل جنس؟

٢) نريد اختيار 5 عدائين من بين 10 عدائين ليشكلوا مجموعة أولى A ويشكل الباقون مجموعة B بكم طريقة يمكن القيام بذلك؟ إذا تضمن العداؤون العشرة ثلاثة سعوديين فما احتمالات أن يكون :

ا ـ كل العدائين السعوديين في المجموعة A؟

ب ـ كل العدائين السعوديين في المجموعة نفسها؟
جـ ـ كل مجموعة تتضمن على الأقل عداء سعوديا؟

٣) من صندوق يتضمن 7 قطع معيبة و8 قطع مقبولة، سحبنا عينة عشوائية من خمس قطع، ما احتمال أن تتضمن العينة:

ا ـ قطعة مقبولة واحدة على الأقل؟
ب ـ قطعا من النوعين؟
جـ ـ إذا كانت قيمة كل قطعة مقبولة 15 ريالا وكانت كل قطعة معيبة تسبب خسارة 5 ريالات، فما هي القيمة المتوقعة للعينة؟ وما هو الانحراف المعياري للمتغير الذي يعبر عن قيمة العينة؟
د ـ أعد حل السؤالين (ا) و(ب) بفرض أن السحب مع الإعادة.

٤) من صندوق يتضمن 7 مقاومات ومقاومة كل منها أوم واحد، وثلاث مقاومات، مقاومة كل منها 4 أوم. اخترنا عشوائيا 4 مقاومات ووصلناها على التسلسل لتشكيل وحدة A، وكذلك وصلنا المقاومات الباقية على التسلسل لتشكيل وحدة B. والمقاومة الكلية لوحدة تساوي مجموع مقاوماتها. ما احتمالات:
ا ـ أن تتضمن كل وحدة مقاومة واحدة على الأقل من نوع الـ 4 أوم؟
ب ـ المقاومة الكلية للوحدة B تتجاوز المقاومة الكلية للوحدة A؟
جـ ـ ما المقاومة الكلية المتوقعة لكل من الوحدتين A، B؟

٥) عدد الأسماك N في حوض للأسماك مقدار غير معروف. اصطدنا عشرين سمكة من الحوض ووضعنا على كل منها علامة مميزة ثم أعدناها إلى الحوض. ثم عدنا فاصطدنا 25 سمكة ووجدنا أنها تتضمن 3 سمكات معلّمة. عبر بدلالة N عن احتمال هذه الحادثة، ولنرمز لهذا الاحتمال بـ P_N. ويعتبر تقديرا جيدا لـ N، تلك القيمة التي تجعل P_N أكبر ما يمكن. وبملاحظة أن:

$$\frac{P_N}{P_{N-1}} = \frac{N^2 - 45N + 500}{N^2 - 42N}$$

نجد أن $P_{N-1} < P_N$ إذا، وفقط إذا، كان $N > 500/3$. ما هو تقديرك لعدد الأسماك في الحوض؟

٦) خذ عينة حجمها 4 من المجتمع المذكور في المثال (٤ ـــ ١٣) واكتب توزيع $\bar{X}$ ، ثم احسب $E(\bar{X})$ و $V(\bar{X})$. ما أثر زيادة حجم العينة على تباين $\bar{X}$ وعلى شكل المدرج الاحتمالي لتوزيع $\bar{X}$ ؟

٧) من بين 15 متقدما لوظيفة ما، تسعة منهم يحملون درجة جامعية. اختير متقدمان عشوائيا لإجراء مقابلة. أوجد احتمال أن:

ا ـ أحدهم يحمل درجة جامعية والآخر لا يحمل درجة جامعية.

ب ـ ليس بينهما من يحمل درجة جامعية.

جـ ـ كلاهما يحمل درجة جامعية.

٨) لدى سكرتير 9 رسائل وعليه أن يرسل ثلاثا منها محددة بالبريد المسجل والباقي بالبريد العادي. اختلط عليه الأمر فاختار عشوائيا 3 رسائل ووضع عليها طوابع البريد المسجل. ما هو احتمال:

ا ـ أن اختياره لم يكن صحيحا تماما،

ب ـ أن اختياره كان صحيحا.

٩) بالاشارة إلى التمرين ١٣ من مجموعة التمارين (١ ـ ١) ، لنعتبر أن القياسات الخمسمائة تشكل مجتمعا من قياسات معدل الكولسترول في الدم. ولنسجل الأعداد الخمسمائة (إذا أمكن)، أو ما يشير إليها، على قطع صغيرة من الورق، ولنختر عشوائيا عينة حجمها عشرة باختيارنا عشوائيا لعشرة أوراق (مع الاعادة). لنكرر تجربة سحب العينة هذه مائة مرة ولنحسب المتوسطات المائة لهذه العينات ونرسم لها مدرج تكرار نسبي مستخدمين حدود الفئات نفسها. قارن الآن مدرج التكرار النسبي الحاصل مع المدرج الخاص بالمجتمع المطلوب في الجزء جـ من التمرين ١٣ من مجموعة التمارين (١ ـ ١). هل تجده أقرب إلى مدرج المجتمع من

المدرجات المطلوبة في الجزء ب من التمرين ١٣ . احسب متوسط البيان الاحصائي من مائة متـوسط وتباينه وقـارنهما مع متوسط المجتمع وتباينـه . ماذا تستنتج؟ (من المفضل أن يتعاون الفصل بكامله في حل هذا التمرين) .