المحاضرة 5: تقارب بعض التوزيعات الاحتمالية

- نتناول في هذا المحور الموسوم بعنوان "تقارب بعض التوزيعات الاحتمالية" بعض حالات التقارب التي تحصل بين عدد من التوزيعات الاحتمالية الشهيرة.
- يقصد بتقارب توزيعين (مثلا ذي الحدين والتوزيع الطبيعي، أو ذي الحدين وتوزيع بواسون...) أن يعطي التوزيعان نتائج متقاربة بخصوص احتمال معين. بمعنى هناك إمكانية لاستخدام توزيعين أو أكثر في حساب احتمال معين.
- فيمايلي بعض من حالات التقارب التي تحدث بين التوزيعات الاحتمالية الشهيرة:

1- تقارب توزيع ذي الحدين (Binomial) إلى التوزيع الطبيعي (Normal):

سؤال: كيف يتم استخدام التوزيع الطبيعي بدلا من توزيع ذي الحدين؟ بعبارة أخرى كيف يتم حساب الاحتمال عن طريق توزيع مستمر بينما المتغيرة العشوائية متقطعة؟

حسب نظرية النهاية المركزية بصيغة موافر ولابلاس Laplace–Moiver Theorem فإنه:

إذا كانت Y_n مجموعة متغيرات عشوائية مستقلة ولكل منها توزيع برنولي أي $Y \sim B(1,p)$ فإن توزيع المجموع $X_n = \sum_1^n Y_i$ الذي يتبع توزيع ذي الحدين $X_n \sim B(n,p)$ سوف يتقارب من التوزيع الطبيعي $N(np, npq)$ كلما كان n كبيرا، أي:

$$\lim_{n\to\infty} B(\mathrm{n},\mathrm{p}) = N(np, npq)$$

حيث:

$$\boldsymbol{\mu = np}$$
$$\boldsymbol{\sigma^2 = npq}$$

- وحتى تكون n كبيرة بما فيه الكفاية، يجب أن تستوفي الشرطين التاليين:

$$np \geq 5$$
$$nq \geq 5$$

- عند توفر هذين الشرطين معا، يمكننا استخدام التوزيع الطبيعي للإجابة على الأسئلة الاحتمالية المتعلقة بالتوزيع ذي الحدين.
- وحتى نتمكن من حساب الاحتمالات للمتغيرة العشوائية X التي تتبع توزيع ذي الحدين، نقوم بتحويلها إلى متغيرة عشوائية أخرى Z تتبع التوزيع الطبيعي المعياري من خلال إجراء التحويل التالي:

$$Z = \frac{X - np}{\sqrt{npq}} \sim N(0\,,1)$$

- وبما أن التوزيع الطبيعي هو توزيع احتمالي مستمر، في حين أن توزيع ذي الحدين هو توزيع احتمالي متقطع، فإنه يجب علينا **تطبيق مفهوم تصحيح الاستمرارية** عند حساب الاحتمالات.
- **تصحيح الاستمرارية** يعني إضافة أو طرح 0.5 إلى قيمة x المتقطعة.

مثال (01):

في تجربة إلقاء قطعة نقود متجانسة 100 مرة، لنفرض أننا نريد إيجاد احتمال الحصول على الصورة 45 مرة. أي إيجاد $P(X = 45)$ في تجربة ذي الحدين. ولاستخدام التوزيع الطبيعي لتقريب ذي الحدين، سنجد بدلا من ذلك $P(44.5 < X < 45.5)$.

الجدول التالي يوضح متى يجب علينا إضافة أو طرح 0.5 لقيمة x، بناء على نوع الاحتمال الذي تريد الحصول عليه:

باستخدام توزيع ذي الحدين	باستخدام التوزيع الطبيعي مع تصحيح الاستمرارية
$X = 45$	$45.5 > X > 44.5$
$X \leq 45$	$X < 45.5$
$X < 45$	$X < 44.5$
$X \geq 45$	$X > 44.5$
$X > 45$	$X > 45.5$

نأخذ مثال آخر لتوضيح خطوات كيفية استخدام التوزيع الطبيعي لتقريب توزيع ذي الحدين:

مثال (02):

في تجربة القاء قطعة نقود متجانسة في الهواء 100 مرة، لنفترض أننا نريد حساب احتمال الحصول على الصورة 45 مرة على الأكثر.

لاحظ أنه في هذا المثال لدينا القيم التالية:

$n = 100$ (عدد مرات رمي قطعة النقود في الهواء).

$X = 45$ (عدد حالات النجاح، أي عدد مرات الحصول على الصورة).

$p = 0.5$ (احتمال الحصول على الصورة، أي احتمال النجاح في الحصول على صورة).

لحساب احتمال الحصول على الصورة 45 مرة على الأكثر نتبع الخطوات التالية:

✓ **الخطوة 01: التأكد من أن حجم العينة كبير بما فيه الكفاية لاستخدام التقريب الطبيعي.**

نتأكد من تحقق الشرطين التاليين:

$$np \geq 5$$
$$nq \geq 5$$

في حالتنا هذه:

$$np = 100(0.5) = 50$$
$$nq = 100(0.5) = 50$$

الشرطين محققين، وبذلك يمكن تطبيق التقريب الطبيعي.

✓ **الخطوة 02: تحديد تصحيح الاستمرارية.**

بالرجوع إلى الجدول أعلاه، نجد أنه يجب علينا إضافة 0.5 عندما نتعامل مع احتمال من الشكل $X \leq 45$، وبالتالي المطلوب منا إيجاد $P(X < 45.5)$.

✓ **الخطوة 03: إيجاد المتوسط μ والتباين σ^2 لتوزيع ذي الحدين.**

$$\boldsymbol{\mu = np} = 100(0.5) = \mathbf{50}$$
$$\boldsymbol{\sigma^2 = npq} = 100(0.5)(0.5) = \mathbf{25}$$

✓ الخطوة **04: تحديد الاحصاءة Z.**

$$\boldsymbol{Z} = \frac{X - np}{\sqrt{npq}} = \frac{45 - 50}{\sqrt{25}} = \mathbf{-1}$$

✓ **الخطوة 05: إيجاد قيمة الاحتمال.**

باستخدام جدول التوزيع الطبيعي المعياري نجد القيمة الاحتمالية المرافقة للقيمة -1 هي 0.1586.

وبالتالي احتمال الحصول على الصورة 45 مرة على الأكثر عند تكرار تجربة رمي قطعة النقود 100 مرة هي 0.1586

2- تقارب توزيع ذي الحدين (Binomial) إلى توزيع بواسون (Poisson):

يعتبر التوزيعان ذي الحدين وبواسون متشابهان في الاحصائيات. قبل تقديم قاعدة التقارب بين التوزيعين، نقدم أولا أوجه التشابه والاختلاف بينهما:

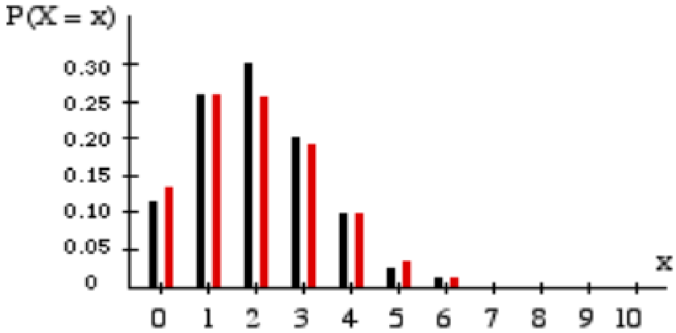
- يمكن استخدام التوزيعان لنمذجة عدد تكرارات بعض الأحداث.
- في كلا التوزيعين، يفترض أن تكون الأحداث مستقلة.

ويكمن الاختلاف بين التوزيعات فيما يلي:

- في توزيع ذي الحدين، هناك عدد محدد من المحاولات (مثلا رمي قطعة نقود 20 مرة).
- في توزيع بواسون، لا يوجد عدد محدد من الأحداث خلال فترة زمنية (مثلا عدد الزبائن الذين يصلون إلى مكتب بريد الجزائر).

بالنسبة للتقارب بين التوزيعين:

- عندما تكون قيمة n في التوزيع ذي الحدين كبيرة وقيمة p صغيرة جدا، يمكن تقريب التوزيع ذي الحدين عن طريق توزيع بواسون.
- إذا كانت $n > 20$ و $np < 5$ أو $nq < 5$، فإن بواسون هو تقدير تقريبي جيد. والتمثيل البياني التالي يوضح أوجه التشابه والتقارب بين التوزيعين، حيث الشكل الأسود يمثل توزيع ذي الحدين عند $n = 10$ و$p = 0.2$. والشكل الأحمر يمثل توزيع بواسون بالمعلمة $\boldsymbol{\lambda} = np = \mathbf{2}$:



- بالنسبة للقيم الكبيرة لـ n والقيم الصغيرة لـ p، فإن توزيع بواسون يقارب التوزيع ذي الحدين.

مثال (03):

يقوم أحد المصانع بوضع البسكويت في صناديق تحتوي كل منها على 100 قطعة. احتمال كسر البسكويت هو 0.03. أوجد احتمال أن يحتوي الصندوق على قطعتين (2) من البسكويت المكسور.

الحل:

- يعتبر هذا التوزيع توزيع ذي الحدين، حيث $n = 100$ و$p = 0.03$

$$\boldsymbol{P(X = 2)} = C_{100}^{2}(0.03)^2(0.97)^{98} = \mathbf{0.2251}$$

- باستخدام تقريب توزيع بواسون:

$$\boldsymbol{\lambda} = np = 100(0.03) = \mathbf{3} < 5$$

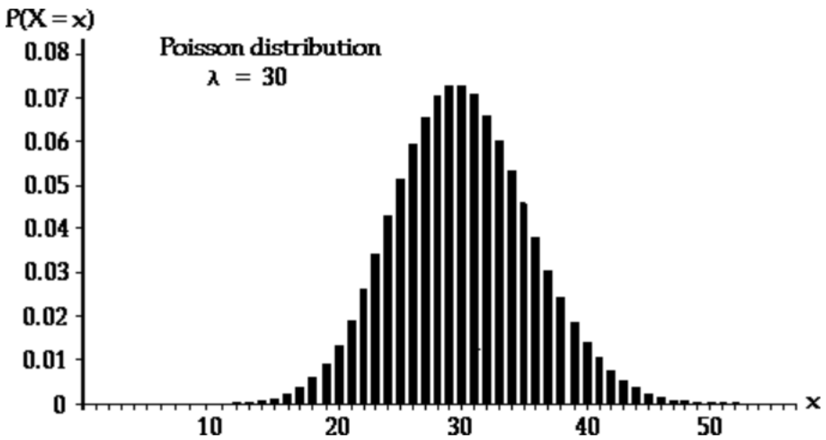
$$\boldsymbol{P(X = 2)} = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{3^2 e^{-3}}{2!} = \mathbf{0.224}$$

وبالتالي احتمال أن يحتوي صندوق البسكويت على قطعتين مكسورتين هو 0.224

3- تقارب التوزيع الطبيعي (Normal) من توزيع بواسون (Poisson):

قبل تقديم قاعدة التقارب بين التوزيعين، نقدم أولا أوجه التشابه والاختلاف بينهما:

- يعد توزيع بواسون والتوزيع الطبيعي من التوزيعات الاحتمالية الأكثر استخداما في الإحصاء.
- توزيع بواسون يصف احتمالية الحصول على x نجاحات خلال فترة زمنية معينة.
- التوزيع الطبيعي يصف احتمال أن يأخذ المتغير العشوائي قيمة خلال فترة زمنية معينة.
- الفرق الأول بين توزيع بواسون والتوزيع الطبيعي هو نوع البيانات التي يستخدمها كل نموذج للتوزيع الاحتمالي.
- يتم استخدام توزيع بواسون عندما نتعامل مع بيانات متقطعة (منفصلة) يمكن أن تأخذ فقط قيما صحيحة تساوي 0 أو 1 أو 2 وهكذا. من بين الأمثلة: عدد المكالمات الواردة في الساعة في مركز الاتصال، عدد الزبائن يوميا في المطعم، عدد حوادث السيارات شهريا.
- يتم استخدام التوزيع الطبيعي عند التعامل مع البيانات المستمرة التي يمكن أن تأخذ أي قيمة من اللانهاية السالبة إلى اللانهاية الموجبة. من بين الأمثلة: وزن حيوان معين، طول نبات معين، درجة الحرارة المئوية
- يمكن تقريب التوزيع الطبيعي إلى توزيع بواسون عندما تكون $\boldsymbol{\lambda > 20}$.
- كلما كبرت λ، يبدأ توزيع بواسون من الاقتراب من التوزيع الاحتمالي الطبيعي. كما هو موضح في الشكل التالي:



- وبما أن التوزيع الطبيعي هو توزيع احتمالي مستمر، في حين أن توزيع بواسون هو توزيع احتمالي متقطع، فإنه يجب علينا **<u>تطبيق مفهوم تصحيح الاستمرارية</u>** عند حساب الاحتمالات.

مثال (04):

إذا كان معدل عدد الحوادث في أحد المصانع هو 45 حادثًا سنويا، وعدد الحوادث سنويا يتبع توزيع بواسون.

استخدم التقريب الطبيعي لإيجاد احتمال وقوع أكثر من 50 حادثًا في السنة.

الحل:

$$\boldsymbol{\mu = \lambda = 45 > 20}$$
$$\boldsymbol{\sigma^2 = \lambda = 45}$$

الشرط $\boldsymbol{\lambda > 20}$ محقق، وبالتالي يمكننا استخدام التقريب الطبيعي في حساب الاحتمالات.

لتكن X هي المتغير العشوائي لعدد الحوادث في السنة.

للحصول على قيمة الاحتمال المطلوب، نقوم أولا بتطبيق تصحيح الاستمرارية بحيث نضيف 0.5 عندما نتعامل مع احتمال من الشكل $X > 50$، بمعنى الاحتمال المطلوب هو $P(X > 50.5)$.

$$P(X > 50.5) = P\left(Z > \frac{50.5 - 45}{\sqrt{45}}\right) = P(Z > 0.820) = 1 - P(Z \leq 0.820) = \mathbf{0.2061}$$