الإحصاء و الإحتمالات

الجزء II: الإحتمالات (Probability)

الفصل 4:المتغيرات العشوائية (random variables)

يتمخّض عن التجربة العشوائية مجموعة من الأحداث الأولية والتي ينجم عنها نتائج إما أن تكون رقمية كما هو الحال في تجربة رمي زهرة نرد، أو أن تكون ترميزية كما هو الحال في تجربة رمي قطعة النقود، فإننا نلجأ إلى إعطاء وصف رياضي أشمل لها بالاعتماد على مصطلح **المتحوّل** أو **المتغير العشوائي**. ونظرا لأهمية المتغيرات العشوائية في تبسيط دراسة التجارب العشوائية فإننا سوف نتطرق في هذا الفصل إلى التعريف بالمتغير العشوائي وأنواعه، القوانين المعتمدة ودوال الاحتمال وكذا التوزيعات والدوال التراكمية، إلى جانب مختلف الخصائص الإحصائية للمتغيرات العشوائية.

1.4.II: المتغير العشوائي (random variable)

لنفترض أننا قمنا بتعيين رقم لكل حالة من الحالات الممكنة من فضاء العينة العشوائية، أي تقديم وصف عددي لنتائج التجربة العشوائية، وبالتالي لدينا دالة محددة في فضاء العينة العشوائية. تسمى هذه الوظيفة بالمتغير العشوائي أو بشكل أكثر دقة دالة عشوائية. ويشار إليه عادة بحرف كبير مثل X أو Y، وبالنسبة للقيم التي يأخذها بأحرف صغيرة $(x_1 , x_2 \ \dots \ x_n)$ أو $(y_1 , y_2 \ \dots \ y_n)$ بحسب الرمز المعبر عن المتغير العشوائي، بينما نرمز للاحتمال الموافق لقيمة المتغير بـــ $P(X = x_i)$ أو $P(Y = y_i)$.

2.4.II: أنواع المتغير العشوائي: تنقسم المتغيرات العشوائية إلى قسمين:

أ/ متغير عشوائي متقطع (منفصل) (Discrete Random Variable): تكون قيمه منفصلة أي يأخذ قيمة من قيم الأعداد الصحيحة والتي تعبّر عن وحدة قياس واحدة وغير قابلة للتجزئة، مثل عدد الذكور في عائلة مكوّنة من 3 أولاد X، $X = \{0,1,2,3\}$

ب/ متغير عشوائي مستمر(متصل) (Continuous Random Variable): تكون قيمه مستمرة في شكل مجال أو عدّة مجالات وبالتالي فهو يعتمد على وحدة قياس قابلة للتجزئة مثل الطول، الوقت، المسافة والوزن.

3.4.II: التوزيع الاحتمالي والتراكمي للمتغير العشوائي المنفصل

1.3.4.II: التوزيع الاحتمالي (Discrete Probability distribution)

إذا كان X متغير عشوائي منفصل، نعتبر أن القيم المحتملة و التي يأخذها معطاة بواسطة $(x_1 , x_2 \ \dots \ x_n)$، كما أن احتمال ظهور أي قيمة x_i هو $P(x_i)$، وبالتالي فإن قانون التوزيع الاحتمالي للمتغيّر العشوائي X يكتب في الجدول التالي:

X	x_1	x_2	x_3	…	x_n	$\sum$
$P(x_i)$	$P(x_1)$	$P(x_2)$	$P(x_3)$	…	$P(x_n)$	1

كما يتطلب أن يتوفر التوزيع الاحتمالي على الخواص الآتية:

$$\begin{cases} 0 \leq \mathrm{P}(x_i) \leq 1 \\ \sum_{i=1}^{n} \mathrm{P}(x_i) = 1 \end{cases}$$

مثال:

رمي قطعة نقود مرتين، نفرض ان X هو عدد الرؤوس.

_ أوجد القانون الاحتمالي للمتغير العشوائي X ؟

الحل:

فضاء العينة العشوائية هو: $\Omega = \{HH, HT, TH, TT\}$

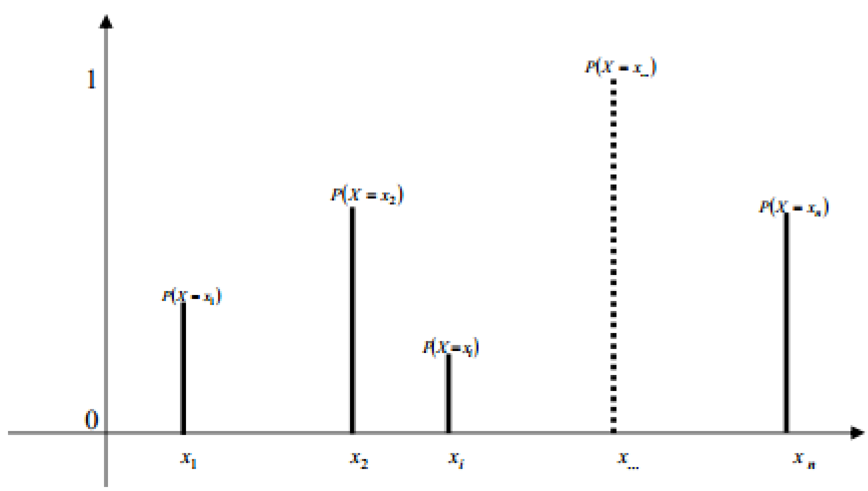
القيم الممكن أن يأخذها: X = 0,1,2

يمكن وضع قانون توزيع الاحتمالات لهذا المتغير العشوائي في الجدول التالي:

X	0	1	2	$\sum$
$\mathrm{P}(X = x_i)$	1/4	2/4	1/4	1

من خلال الجدول نلاحظ أن $\sum_{i=1}^{n} \mathrm{P}(x_i) = 1$، وكل قيمة متعلقة بالاحتمال محصورة بين الصفر و الواحد.

● **التمثيل البياني لقانون الاحتمال:** يأخذ التعبير الهندسي للعلاقة بين قيم المتغير العشوائي وقيم الاحتمال المقابلة لها الشكل البياني كما يلي:



II.2.3.4: **التوزيع التراكمي** (Distribution Functions or cumulative distribution)

تقيس دالة التوزيع التراكمي (cdf)، المشار إليها بالرمز $F(x)$، احتمال أن يأخذ المتغير العشوائي X قيمة أقل من أو تساوي x، أي:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(x_i)$$

كما يمكن التعبير عن التوزيع التراكمي بأسلوب مماثل لقانون الاحتمال، ويتم ذلك إما بالاعتماد على الجدول أو الشكل البياني، وذلك على النحو الأتي:

X	x_1	x_2	x_3	…	x_n	$\sum$
$P(x_i)$	$P(x_1)$	$P(x_2)$	$P(x_3)$	…	$P(x_n)$	1
$F(x)$	$P(x_1)$	$P(X \leq x_2)$	$P(X \leq x_3)$	…	$P(X \leq x_n)$	–

مثال: بالرجوع للمثال السابق :

_أوجد دالة التوزيع $F(x)$ ؟

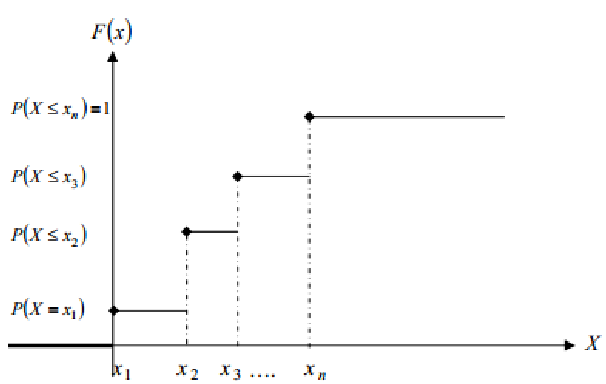
_ما هو احتمال أن يكون عدد الرؤوس الظاهرة لا يتعدى 1؟

الحل:

X	0	1	2	$\sum$
$P(X = x_i)$	1/4	2/4	1/4	1
$F(x)$	1/4	3/4	1	–

_ احتمال أن يكون عدد الرؤوس الظاهرة لا يتعدى 1 هو: $F(x) = P(X \leq x_2) = P(x_1) + P(x_2) = 3/4$

- **التمثيل البيـاني للتوزيع التراكمي**: يمكن تمثيل التوزيع التراكمي للمتغير العشوائي المنفصل بيانيا كما يلي:



4.4.II: دالة الكثافة الاحتمالية والتراكمية للمتغير العشوائي المتصل

1.4.4.II: دالة الكثافة الإحتمالية (Probability Density Function)

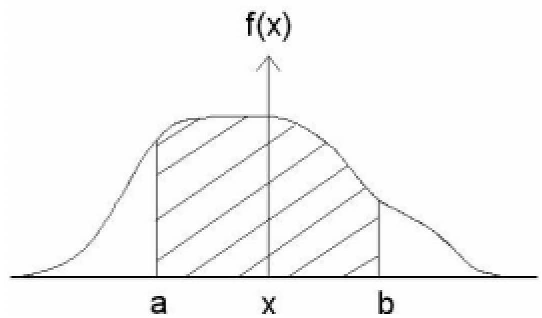
في حالة X متغير عشوائي مستمر فإن قانون التوزيع الإحتمالي يكون في شكل تابع حقيقي (دالة) معرف على R، يدعى بـ**دالة الكثافة الإحتمالية** ويرمز لها بالرمز $f(x)$، وحتى يكون X قانون توزيع احتمال يجب أن تحقق دالة الكثافة الخاصيتين التاليتين:

$$\begin{cases} f(x) \geq 0 \\ \int_{-\infty}^{+\infty} f(x)\,dx = 1 \end{cases}$$

وعليه فإن احتمال X محصور بين عددين a و b بحيث $b > a$ يعطي كمايلي:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

ويتم التعبير عن هذه القيمة الاحتمالية هندسيا كيما يلي:



حيث تساوي مساحة الحيز المحصور بين المنحنى $f(x)$ وخط المحور X في حدود الفاصلتين a و b.

ملاحظات:

1_ $\int_a^a f(x)\,dx = 0$

2_ $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

2.4.4.II: دالة التوزيع للمتغير العشوائي المستمر (Cumulative Distribution Function)

هي مقدار التكامل الرياضي لدالة الكثافة الاحتمالية ضمن المجال $]-\infty, x[$، وتعطى بالصيغة التالية:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

ويتم التعبير عن دالة التوزيع التراكمية بالشكل الأتي:

$$F(x)=\begin{cases}0 & X<a\\ \int_{-\infty}^{x} f(u)\,du & a\le X\le b\\ 1 & X\ge b\end{cases}$$

- أهم خواص دالة التوزيع التراكمية للمتغير العشوائي المستمر:

  - $1\ge F(x)\ge 0$
  - $F(+\infty)=1\ ;\ \ F(-\infty)=0$
  - $\forall(a,b)\in\Re:\ \ P(a\le x\le b)=\int_a^b f(x)dx=F(b)-F(a)$
  - $P(X\le x)=P(X<x)$
  - $P(a<X<b)=\int_a^b f(x)dx$
  - $P(X<x)=1-P(X\ge x)$
  - $P(X<b)=P(X\le b)=\int_{-\infty}^{b} f(x)dx=F(b)$
  - $P(X>b)=P(X\ge b)=\int_{b}^{+\infty} f(x)dx=1-F(b)$
  - $P(X<x)=1-P(X\ge x)$

مثال:

إذا كانت المدة الزمنية التي يستغرقها الطالب في إتمام الاختبار (المدة القانونية ساعتين)، تمثل متغير عشوائي مستمر X (بالساعات)، ودالة الكثافة الاحتمالية معرفة كما يلي:

$$f(x)=\begin{cases}x^2+\frac{4}{3}x\ ;if\ \ 0\le X\le 1\\ 0;\ \text{Other wise}\end{cases}$$

المطلوب:

-تحقق من أن التابع $f(x)$ عبارة عن دالة كثافة احتمالية؟

-تم اختيار طالب عشوائيا، ما هو احتمال أن ينجز الاختبار في أقل من نصف ساعة؟

الحل:

_ التحقق من أن $\mathrm{f}(x)$ دالة كثافة احتمالية:

نلاحظ أن: $\mathrm{f}(x) \geq 0 \quad if \quad 0 \leq X \leq 1$

ومنه الشرط الأول محقق.

حساب التكامل: $\int_{-\infty}^{+\infty} \mathrm{f}(x)\, dx$

$$\int_{-\infty}^{+\infty} \mathrm{f}(x)\, dx = \int_{-\infty}^{0} \mathrm{f}(x)\, dx + \int_{0}^{1} \mathrm{f}(x)\, dx + \int_{1}^{+\infty} \mathrm{f}(x)\, dx$$

$$= 0 + \int_0^1 \left(x^2 + \frac{4}{3}x\right) dx + 0$$

$$= \int_0^1 x^2\, dx + \frac{4}{3}\int_0^1 x\, dx \ = \left[\frac{x^3}{3}\right]_0^1 + \frac{4}{3}\left[\frac{x^2}{2}\right]_0^1$$

$$= \frac{1^3}{3} + \frac{4}{3} \times \frac{1^2}{2} = 1$$

ومنه الشرط الثاني محقق، وبالتالي $\mathrm{f}(x)$ دالة كثافة احتمالية.

_ حساب احتمال أن ينجز الطالب الذي تم اختياره الإختبار في أقل من نصف ساعة:

$$\mathrm{P}\left(X \leq \frac{1}{2}\right) = \int_0^{1/2} \mathrm{f}(x)\, dx = \int_0^{1/2} \left(x^2 + \frac{4}{3}x\right) dx$$

$$= \int_0^{1/2} x^2\, dx + \frac{4}{3}\int_0^{1/2} x\, dx$$

$$= \left[\frac{x^3}{3}\right]_0^{1/2} + \frac{4}{3}\left[\frac{x^2}{2}\right]_0^{1/2} = \frac{(1/2)^3}{3} + \frac{4}{3} \times \frac{(1/2)^2}{2} = 0.2$$

أي أن احتمال ان تكون المدة التي يستغرقها طالب تم اختياره عشوائيا في إنجاز الاختبار أقل من نصف ساعة هي: 0.2 ساعة. من جهة أخرى هذا يشير إلى أن حوالي 20% من الطلبة يمكنهم أن ينجزو الاختبار في أقل من نصف ساعة.

II. 5.4: الخصائص الإحصائية للمتغير العشوائي

II.1.5.4: التوقع الرياضي (mean or expected value):

وهو عبارة عن متوسط المتغير العشوائي الذي يحدد القيمة التي تتمركز حولها باقي القيم التي يأخذها المتغير، يشار إليه بالرمز $\mathrm{E}(x)$ او μ. ويحسب التوقع الرياضي حسب طبيعة المتغير العشوائي، لذلك نميز حالتين:

- **حالة المتغير العشوائي المتقطع:**

إذا كان X متغير عشوائي منفصل، نعتبر أن القيم المحتملة و التي يأخذها معطاة بواسطة $(x_1, x_2 \dots x_n)$، كما أن احتمال ظهور أي قيمة x_i هو $P(x_i)$، وبالتالي فإن التوقع الرياضي للمتغير العشوائي المتقطع يعطى كما يلي:

$$E(x) = \sum_{i=1}^{n} P(x_i)x_i$$

- **حالة المتغير العشوائي المستمر:**

اذا كان المتغير العشوائي X مستمر وله دالة كثافة إحتمالية $f(x)$، فإن الصيغة الرياضية للتوقع الرياضي تعطى كما يلي:

$$E(x) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

II.4.5.2: التباين (variance)

يرمز له بالرمز $V(x)$ أو σ^2، يعبر عن مقدار مدى تشتت قيم المتغير العشوائي حول توقعها الرياضي، ويعرف أيضا بالعزم المركزي من الدرجة الثانية، ياخد نفس الصيغة سواءا مع متغير عشوائي متقطع او مستمر ويعطى بالعبارة التالية:

$$V(x) = E(x^2) - [E(x)]^2$$

حيث :

$$E(x^2) = \int_{-\infty}^{+\infty} x^2\, f(x)\, dx$$

II.4.5.3: الانحراف المعياري (standard deviation)

يرمز له بالرمز σ، يقصد به المتوسط التربيعي للانحرافات عن التوقع الرياضي لقيم المتغير العشوائي ويعطى بالعلاقة التالية:

$$\sigma = \sqrt{V(x)}$$

مثال:

ليكن X متغيرا عشوائيا ويتم التعبير عن تغيراته بالدالة الإحتمالية التالية:

$$f(x) = \begin{cases} \frac{1}{2}x\,; & if\;\; 0 \le X \le 2 \\ 0; & \text{Other wise} \end{cases}$$

المطلوب:

1_ أرسم دالة الكثافة الإحتمالية لهذا المتغير العشوائي

2_ أحسب التوقع الرياضي $E(x)$

3_ أحسب الانحراف المعياري

4_ أوجد الإحتمالات التالية: $P(1 < X < 2)$، $P(X < 1)$، $P(X \ge 1)$، $P(X = 1)$.