الفصل الخامس: المتغيرات العشوائية الثنائية:

المتغيرة الثنائية الاستقلال التباين و الارتباط

1 - المتغيرة الثنائية

1 التوزيعات المشتركة المتقطعة والدالة الهامشية (الحدية) Fonction marginale

(أ) **تعريف:** المتغيرة الثنائية هي متغيرة تتوقف ليس على قيمة واحدة هي قيمة X مثلا و إنما تتوقف على قيمة متغيرتين اثنتين. مثال ذلك، معدل الطالب يتوقف على نقطة الرقابة المستمرة و نقطة التطبيق أو نقطة السداسي الأول ونقطة السداسي الثاني. كذلك نتيجة السنة المالية تتوقف على متغيرتي التكاليف و الإيرادات، وهكذا. التعريف الدقيق للمتغيرة الثنائية يتأتى باستخدام الترميز كما يلي:

لتكن لدينا متغيرتان عشوائيتان متقطعتان X و Y، لنرمز للاحتمال: $P(X = x, Y = y)$ ب $f(x,y)$:

$$f(x,y) = P(X = x, Y = y)$$
$$f(x,y) \geq 0$$
$$\sum_x \sum_y f(x,y) = 1$$

تسمى الثنائية (X,Y) متغيرة ذات بعدين و$f(x,y)$ دالة الكثافة الاحتمالية لها ونقول أيضا دالة الكثافة المشتركة للمتغيرتين X و Y ويمكن التعبير عنها عن طريق جدول للاحتمالات المشتركة (جدول التوزيع المشترك).

Y / X	y_1	y_2	. . .	y_m	$f_1(x)$
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	. . .	$f(x_1,y_m)$ →	$f_1(x_1)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	. . .	$f(x_2,y_m)$	$f_1(x_2)$
. . .	. . .	. . .	. . .	. . .	. . .
x_n	$f(x_n,y_1)$ ↓	$f(x_n,y_2)$	. . .	$f(x_n,y_m)$	$f_1(x_n)$
$f_2(y)$	$f_2(y_1)$	$f_2(y_2)$	. . .	$f_2(y_n)$	1

احتمال $X = x$ يحسب ويكتب كما يلي $P(X = x) = f_1(x) = \sum_{k=1}^{m} f(x,y_k)$

احتمال $Y = y$ يحسب و يكتب كما يلي $P(Y = y) = f_2(y) = \sum_{i=1}^{n} f(x_i,y)$

الدالتين $f_1(x)$ و$f_2(y)$ تسميان الدالتان الهامشيتان (الحديتان) حيث : $\sum f_1(x) = 1$ و $\sum f_2(y) = 1$

(ب) الدالة التجميعية

الدالة التجميعية للمتغيرة الثنائية (X,Y) تكتب كما يلي:

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v)$$

مثال: نرمي قطعة نقدية وحجر نرد، نرمز بX لعدد مرات ظهور الصورة، وY للرقم الذي يظهر من مكعب النرد.

- أكتب التوزيع الاحتمالي المشترك للمتغيرتين،
- أحسب احتمالات الأحداث التالية: الحصول على صورة مع الرقم 6، الحصول على الصورة، الحصول على الرقم 6.
- أحسب الاحتمال $P(X \leq 1, Y \leq 3)$ ، $P(X \leq 2, Y \leq 6)$.

X\Y	1	2	3	4	5	6	$f_1(x)$
0	1/12	1/12	1/12	1/12	1/12	1/12	**½**
1	1/12	1/12	1/12	1/12	1/12	**1/12**	**½**
$f_2(y)$	1/6	1/6	1/6	1/6	1/6	1/6	1

احتمال الحصول على صورة مع الرقم 6 : $P(X = 1 \text{ et } Y = 6) = f(1, 6) = 1/12$

احتمال الحصول على الصورة $P(X = 1) = f_1(1) = \sum_{k=1}^{n} f(1, y_k) = 1/12 + 1/12 + \ldots = 1/2$

احتمال الحصول على الرقم 6 $P(Y = 6) = f_2(6) = \sum_{i=1}^{m} f(x_i,6) = 1/12 + 1/12 = 1/6$

$P(X \le 1, Y \le 3) = F(1, 3) = \sum_{u\le1} \sum_{v\le3} f(u,v) = 6(1/12) = 1/2$, $P(X \le 2, Y \le 6) = 1$

2 التوزيعات المشتركة المتصلة

(أ) تعريف

لتكن لدينا X و Yمتغيرتان ع متصلتان، نعرف دالة الكثافة الاحتمالية المشتركة لهما (دالة الكثافة الاحتمالية للثنائية (X,Y)) كما يلي:

$$P(X = x, Y = y) = f(x, y)$$

$$f(x, y) \ge 0$$

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

(ب) الدالة التجميعية

نكتب دالة التوزيع (الدالة التجميعية) كما يلي:

$$F(x, y) = P(X \le x, Y \le y) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{y} f(u, v)\, dv\, du$$

و يمكن استنتاج دالة الكثافة المشتركة من الدالة التراكمية بالاشتقاق كما يلي: $f(x, y) = \partial^2 F / (\partial x\, \partial y)$

$F_1(x) = P(X \le x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} f(u, v)\, dv\, du$ من جهة أخرى

$$F_2(y) = P(Y \le y) = \int_{u=-\infty}^{\infty}\int_{v=-\infty}^{y} f(u, v)\, dv\, du$$

ونسمي الدالتان $F_1(x)$ و$F_2(y)$ الدالتان التراكميتان (التجميعيتان) الهامشيتان (الحديتان).

ولتحديد احتمال X و Y محصورتان في مجالين ما نكتب:

$$P(a \le x \le b, c \le y \le d) = \int_{x=a}^{b}\int_{y=c}^{d} f(x, y)\, dy\, dx$$

(ج) الدوال الهامشية

الدالتان الهامشيتان (الحديتان) للكثافة الاحتمالية للثنائية (X, Y) فيعبر عنها كما يلي:

$$f_1(x) = \int_{v=-\infty}^{+\infty} f(x, v)\, dv\,, \qquad f_2(y) = \int_{u=-\infty}^{+\infty} f(u, y)\, du$$

مثال: لدينا دالة الكثافة الاحتمالية المشتركة للمتغيرتين X و Y المعرفة كما يلي:

$$f(x,y)=\begin{cases}\dfrac{xy}{96} & , \quad 0<x<4,\ 1<y<5\\ 0 & \text{sinon}\end{cases}$$

أكتب دالة التوزيع الهامشية لكل من المتغيرتين أحسب احتمال $0 < x < 2$ ، أحسب احتمال $1 < y < 3$.

$$F_1(x) = P(X \le x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} f(u, v)\, dv\, du$$

* $x < 0 : F_1(x) = 0,$

* $0 \le x < 4 :$

$$F_1(x) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{\infty} uv/96\, du\, dv = 0 + \int_{u=0}^{x}\int_{v=1}^{5} uv/96\, dv\, du$$

$$= 1/96 \int_{u=0}^{x}[\int_{v=1}^{5} uv\, dv]\, du = 1/96\int_{u=0}^{x}[12u]\, du = x^2/2\ (12/96) = \mathbf{x^2/16}.$$

* $x \ge 4: F_1(x) = 1$

$$F_1(x)=\begin{cases}0 & x<0\\ x^2/16 & 0\le x<4\\ 1 & x\ge 4\end{cases}$$

$$F_2(y) = P(Y \le y) = \int_{u=-\infty}^{\infty}\int_{v=-\infty}^{y} f(u, v)\, dv\, du$$

Pour $y < 1 : F_2(y) = 0,$

* $1 \le y < 5 :$

$$F_2(y) = 0 + \int_{u=0}^{4}\int_{v=1}^{y} uv/96\, dv\, du = 1/96 \int_{u=0}^{4}[\int_{v=1}^{y} uv\, dv]\, du = 1/96 \int_{u=0}^{4}[u\,(y^2 - 1) / 2]\, du$$

$$= (1/2 * 1/96)\,(y^2 - 1)\,(u^2/2)_0^4 = (1/(2*96))\,(y^2 - 1)\,(16/2) = (y^2 - 1) / 24$$

* $y \ge 5 : F_2(y) = 1$

$$F_2(y)=\begin{cases}0 & y<1\\ (y^2-1)/24 & 1\le y<5\\ 1 & y\ge 5\end{cases}$$

$P(0 < x < 2) = F_1(2) - F_1(0) = 4/16 = ¼ ,$

$P(1 < y < 3) = 8 / 24 = 1/3$

3 التوزيع الشرطي Distribution conditionnelle

في حالة X، Y متغيرتان عشوائيتان متقطعتان، فإن دالة الكثافة الاحتمالية الشرطية ل (X|Y = y) تكتب كما يلي f(x/y) وتحسب كما يلي:

$$f(x/y) = \frac{f(x,y)}{f_1(x)} = \frac{P(X = x, Y = y)}{P(X = x)}$$

و هذا استنادا إلى القانون التقليدي للاحتمالات الشرطية: $P(B/A) = \dfrac{P(A \cap B)}{P(A)}$

التوزيع الاحتمالي ل X حيث Y = y هو مجموعة قيم المتغيرة X عند تثبيت Y والاحتمالات f(x/y) المقابلة لها.

4 خلاصة

احتمال ثنائية عشوائية و يحسب كما يلي: $f(x,y) = P(X = x, Y = y)$

للتعبير عن احتمال قيمة ما لإحدى المتغيرتين نكتب: $P(X = x) = f_1(x)$ و تسمى دالة الكثافة الهامشية.

الاحتمال التجميعي لمتغيرتين فيعبر عنه من خلال دالة التوزيع المشتركة: $F(x, y) = P(X \leq x, Y \leq y)$

تنطبق هذه التعاريف على كل من المتغيرات المتقطعة و المستمرة.

للتعبير عن التوزيع الاحتمالي لإحدى المتغيرتين بشرط أن تأخذ المتغيرة الثانية قيمة ما (0، مثلا) فنكتب:

$P(X/Y = 0)$

لحساب الاحتمالات الشرطية ل X تستخدم القاعدة: $P(X/Y = y) = \frac{f(x,y)}{f_1(x)} = \frac{P(X = x, Y = y)}{P(X = x)}$

2 - الاستقلال التباين والارتباط

5 تعريف استقلال متغيرتين

لدينا حدثين عشوائيين A وB يكونان مستقلان إذا كان:

$$P(A \text{ et } B) = P(A)\,P(B)$$

انطلاقا من هذه القاعدة، تكون المتغيرتان العشوائيتان المتقطعتان X و Y مستقلتان إذا وفقط إذا كان:

$$P(X = x, Y = y) = P(X = x)\,P(Y = y)$$

$$f(x, y) = f_1(x)\,f_2(y)$$

في حالة كون المتغيرتين متصلتين نكتب: $P(X \leq x, Y \leq y) = P(X \leq x)\,P(Y \leq y)$

$$F(x, y) = F_1(x)\,F_2(y)$$

أي أن المتغيرتان المستقلتان هما اللتان يمكن كتابة دالة التوزيع المشتركة لهما (أو دالة الكثافة المشتركة) في شكل جداء دالتين هامشيتين تراكميتين (أو دالتين هامشيتين للكثافة).

مثال. ليكن X و Y م ع مستمرين حيث دالة الكثافة المشتركة لهما معرفة كما يلي:

$$f(x,y) = \begin{cases} cxy & , \quad 0 < x < 4,\ 1 < y < 5 \\ 0 & \quad \text{sinon} \end{cases}$$

بين أن المتغيرتين X و Y مستقلتين.

Soit $c = c_1 * c_2 \Rightarrow f(x, y) = c_1\,c_2\,xy = c_1x * c_2y \Rightarrow f(x, y) = f_1(x) * f_2(y)$ cqfd

6 توقع وتباين المتغيرة العشوائية متعددة الأبعاد

ينطبق كل من تعريف التوقع الرياضي والتباين فيما سبق على المتغيرة العشوائية متعددة الأبعاد. لتكن X وYمتغيرتان عشوائيتان متقطعتان، و$f(x, y)$ دالة كثافة مشتركة لهما.

$$\mu_y = E(Y) = \sum_x\sum_y y\, f(x, y) \qquad \mu_x = E(X) = \sum_x\sum_y x\, f(x, y) \quad ,$$

$$\sigma^2_x = E[(x - \mu_x)^2] = \sum_x\sum_y (x - \mu_x)^2\, f(x, y)\ ,\ \sigma^2_y = E[(y - \mu_y)^2] = \sum_x\sum_y (y - \mu_y)^2\, f(x, y)$$

في حالة X و Y متغيرتان مستمرتان:

$$\mu_x = E(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, f(x, y)\, dx\, dy\ ,\ \mu_y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\, f(x,y)\, dx\, dy\ .$$

$$\sigma^2_x = E[(x - \mu_x)^2] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - \mu_x)^2\, f(x, y)\, dx\, dy\ ,$$

$$\sigma^2_y = E[(y - \mu_y)^2] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y - \mu_y)^2\, f(x, y)\, dx\ \ dy$$

مثال. ليكن لدينا التوزيع المشترك المعرف كما يلي:

X \ y	-4	-2	7
1	1/8	1/4	1/8
-5	1/4	1/8	1/8

المطلوب حساب:

σ^2_x, σ^2_y ، E(y), E(x)

$E(x) = \sum_x\sum_y x\, f(x, y) = 1(1/8 + ¼ + 1/8) – 5(1/4 + 1/8 + 1/8) = 1/2 – 5/2 = -4/2 = -2$

$E(Y) = \sum_x\sum_y y\, f(x, y) = -4\ (1/8 + ¼) – 2\ (1/4 + 1/8) + 7\ (1/8 + 1/8) = -1/2$

$\sigma^2_x = E[(x – \mu_x)^2] = \sum_x\sum_y (x - \mu_x)^2\, f(x, y)$

$= (1 + 2)^2\ (1/8 + ¼ + 1/8) + (-5 + 2)^2\ (1/4 + 1/8 + 1/8) = 9\ (1/2) + 9\ (1/2) = 9$

$\sigma^2_y = E[(y – \mu_y)^2] = \sum_x\sum_y (y - \mu_y)^2\, f\ (x, y)$

$= (-4 + 1/2)^2\ (1/8 + 1/4) + (-2 + 1/2)^2\ (1/4 + 1/8) + (7 + 1/2)^2\ (1/8 + 1/8)$

$= 49/4\ \ (3/8) +\ 9/2\ (3/8) + (15/2)^2\ (2/8) = 651 / 32 = 20,34$

كما يمكن حساب كل من القيم السابقة باستخدام الدوال الهامشية $f_1(x)$ و $f_2(y)$.

X \ Y	-4	-2	7	**$f_1(x)$**
1	1/8	1/4	1/8	**4/8**
-5	1/4	1/8	1/8	**4/8**
$f_2(y)$	**3/8**	**3/8**	**2/8**	1

$E(x) = \sum_x x\, f_1(x) = 1(4/8) – 5(4/8) = -2$

$V(x) = E(x^2) –(E\ (x))^2.$

$= [1^2(4/8) + (– 5)^2(4/8)] – (–2)^2 = 9$

7 التباين المشترك Covariance

يعرف التباين المشترك كمايلي:

$$Cov\ (X, Y) = (\sigma_{xy})^2 = E[(X – \mu_x)(Y – \mu_y)]$$

في حالة X و Yمتغيرتان متقطعتان:

$$(\sigma_{xy})^2 = \sum_x\sum_y (x – \mu_x)(y – \mu_y) f(x, y)$$

في حالة X و Yمتغيرتان مستمرتان:

$$(\sigma_{xy})^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - \mu_x)\,(y - \mu_y)\,f(x, y)\,dx\,dy$$

(أ) **خصائص التباين المشترك**

1. من تعريف التباين يمكن أن نستنتج:

$$\begin{aligned}Cov(X, Y) &= E[(X - \mu_x)(Y - \mu_y)]\\ &= E[XY - X\mu_y - \mu_x Y + \mu_x\mu_y]\\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + \mu_x\mu_y\\ &= E(XY) - E(X)\,E(Y)\end{aligned}$$

$$Cov(X, Y) = E(XY) - E(X)\,E(Y)$$

2. في حالة X و Y متغيرتان مستقلتان، نعلم من خصائص التوقع الرياضي أن $E(XY) = E(X)\,E(Y)$ ومنه:

$$Cov(X, Y) = E(XY) - E(X)\,E(Y) = E(XY) - E(XY)$$
$$Cov\,(X, Y) = 0$$

العكس ليس بالضرورة دوما صحيح، فقد يكون التباين المشترك مساويا للصفر من غير أن يكون المتغيرتان مستقلتان فعلا. المعادلة هي في الحقيقة تمثل شرطا ولكنه ليس شرطا كافيا. بالمقابل يمكن استعمال نتيجة معدومية التباين المشترك للدلالة على ضعف الارتباط، إذا كان موجودا، بين المتغيرتين.

3. في حالة X و Yمتغيرتان مستقلتان أو غير مستقلتين:

$$Var\,(X \pm Y) = V(X) + V(Y) \pm 2\,Cov\,(X, Y)$$

4. القيمة المطلقة للتباين المشرك لا تكون أكبر من جداء الانحرافين المعياريين: $|(\sigma_{xy})^2| \le \sigma_x\sigma_y$

5. في حالة X و Yمتغيرتان مرتبطتان تماما مثلا $Y = X$ فإن: $Cov(X, Y) = V(X)\,V(Y)$

8 معامل الارتباط

من الخاصية (2) نستنتج أن الكسر $\dfrac{\sigma_{xy}^2}{\sigma_x\sigma_y}$ يساوي 0 في حالة X وY مستقلتان و من الخاصية (5) نستنتج أنه في حالة X و Yمتغيرتان مرتبطتان تماما فإن الكسر يساوي 1.

من جهة أخرى من الخاصية 4 نستنتج أن النسبة تتراوح قيمته بين (−1) و(1): $-1 \le \dfrac{\sigma_{xy}^2}{\sigma_x\sigma_y} \le 1$. من أجل هذا تستعمل النسبة:

$r = \dfrac{\sigma_{xy}^2}{\sigma_x\sigma_y}$ لقياس الارتباط بين المتغيرتين، وتسمى معامل الارتباط.

في حالة r معدوم نقول أن **المتغيرتان غير مرتبطتين**، من غير أن نجزم أنهما مستقلتان.