回帰分析I

7. 二次元の確率変数

2つの確率変数の分布

- ここまでは1つの確率変数について考えました。
- ここでは2つの確率変数の分布について考えていきます。
  - 回帰分析では、二つ以上の確率変数を使ってのモデリング。
- 例を使って説明していきます。
- 二つの確率変数：XとY
  - Xが取りうる値は1と2
  - Yが取りうる値は5と10
- 取りうる値の組み合わせとその組み合わせが起こる確率（**結合確率分布**という）は以下の通り。

$\mathrm{Pr}(\mathrm{X}=1, \mathrm{Y}=5)=0.2$
確率変数Ｘが1、かつ確率変数Ｙが5の値をとる確率

	Y = 5	Y = 10	
X = 1	0.2	0.4	
X = 2	0.1	0.3	

$\mathrm{Pr}(\mathrm{X}=1, \mathrm{Y}=10)=0.4$
確率変数Ｘが1、かつ確率変数Ｙが10の値をとる確率

	Y = 5	Y = 10	
X = 1	0.2	0.4	
X = 2	0.1	0.3	

$\Pr(X = 2, Y = 5) = 0.1$
確率変数Ｘが2、かつ確率変数Ｙが5の値をとる確率

	Y = 5	Y = 10	
X = 1	0.2	0.4	
X = 2	0.1	0.3	

$\Pr(X = 2, Y = 10) = 0.3$
確率変数Ｘが2、かつ確率変数Ｙが10の値をとる確率

	Y = 5	Y = 10	
X = 1	0.2	0.4	
X = 2	0.1	0.3	

一般的には、、、、

- 確率変数Xの取りうる値（実現値）： $\{x_1, \ldots, x_m\}$
- 確率変数Yの取りうる値（実現値）： $\{y_1, \ldots, y_n\}$
- $Pr(X = x_i, Y = y_j) = h(x_i, y_j)$ $(i = 1, \ldots, m; j = 1, \ldots, n)$と表そう。
- このとき$h(x_i, y_j)$を「**結合確率質量関数**」と呼びます。
- 例では $h(1,5) = 0.2;$ $h(1,10) = 0.4;$ $h(2,5) = 0.1;$ $h(2,10) = 0.3$
- $h(x_i, y_j)$ は 確率なので以下の条件を満たします。

  $h(x_i, y_j) \geq 0$ （確率は非負）
  $\sum_{i=1}^{m} \sum_{j=1}^{n} h(x_i, y_j) = 1$ （起こりうる場合の確率全部足すと1）

確率変数X が 1の値をとる確率は？

	Y = 5	Y = 10	
X = 1	0.2	0.4	
X = 2	0.1	0.3	

確率変数X が 1の値をとる確率は？

	Y = 5	Y = 10	
X = 1	0.2	0.4	P(X=1) = 0.6
X = 2	0.1	0.3	

$P(X = 1) = P(X = 1, Y = 5) + P(X = 1, Y = 10) = h(1,5) + h(1,10)$

確率変数X が 2の値をとる確率は？

	Y = 5	Y = 10	
X = 1	0.2	0.4	
X = 2	0.1	0.3	

確率変数X が 2の値をとる確率は？

	Y = 5	Y = 10	
X = 1	0.2	0.4	
X = 2	0.1	0.3	P(X=2) = 0.4

$P(X = 2) = P(X = 2, Y = 5) + P(X = 2, Y = 10) = h(2,5) + h(2,10)$

これはＸだけについての分布。これをＸの**「周辺分布」**と呼びます。

	Y = 5	Y = 10	
X = 1	0.2	0.4	P(X=1) = 0.6
X = 2	0.1	0.3	P(X=2) = 0.4

$P(X = 1) = P(X = 1, Y = 5) + P(X = 1, Y = 10) = h(1,5) + h(1,10)$

$P(X = 2) = P(X = 2, Y = 5) + P(X = 2, Y = 10) = h(2,5) + h(2,10)$

- 一般的に、Xの周辺分布は、XとYの結合分布から次のように求められます。

- XとYの結合確率質量関数:$h(x,y)$
- Xの周辺確率質量関数を$f(x)$とすると

$$f(x) = \sum_{y} h(x,y)$$

- 例だと

$$\Pr(X=1) = f(1) = h(1,5) + h(1,10)$$
$$\Pr(X=2) = f(2) = h(2,5) + h(2,10)$$

だから

$$f(x) = \begin{cases} 0.6 & (x=1) \\ 0.4 & (x=2) \end{cases}$$

確率変数Y が 5の値をとる確率は？

	Y = 5	Y = 10	
X = 1	0.2	0.4	P(X=1) = 0.6
X = 2	0.1	0.3	P(X=2) = 0.4
	P(Y=5) = 0.3		

P(Y = 5) = P(X = 1, Y = 5) + P(X = 2, Y = 5) = h(1,5) + h(2,5)

確率変数Y が 10の値をとる確率は？

	Y = 5	Y = 10	
X = 1	0.2	0.4	P(X=1) = 0.6
X = 2	0.1	0.3	P(X=2) = 0.4
	P(Y=5) = 0.3	P(Y=10) = 0.7	

P(Y = 10) = P(X = 1, Y = 10) + P(X = 2, Y = 10) = h(1,10) + h(2,10)

これはＹだけについての分布。これをＹの**「周辺分布」**と呼びます。

	Y = 5	Y = 10	
X = 1	0.2	0.4	P(X=1) = 0.6
X = 2	0.1	0.3	P(X=2) = 0.4
	P(Y=5) = 0.3	P(Y=10) = 0.7	

P(Y = 5) = P(X = 1, Y = 5) + P(X = 2, Y = 5) = h(1,5) + h(2,5)

P(Y = 10) = P(X = 1, Y = 10) + P(X = 2, Y = 10) = h(1,10) + h(2,10)

- 一般的に、Yの周辺分布は、XとYの結合分布から次のように求められます。

- XとYの結合確率質量関数：$h(x, y)$
- Yの周辺確率質量関数を$g(y)$とすると

$$g(y) = \sum_{x} h(x, y)$$

- 例だと

$$\Pr(Y = 5) = g(5) = h\,(1{,}5) + h\,(2{,}5)$$
$$\Pr(Y = 10) = g(10) = h\,(1{,}10) + h\,(2{,}10)$$

だから

$$g(y) = \begin{cases} 0.3 \ (y = 5) \\ 0.7 \ (y = 10) \end{cases}$$

17

結合確率分布：連続的確率変数の場合

- X、Yがそれぞれ連続的確率変数で結合確率密度関数$h(x, y)$をもつとします。

- このときXの周辺確率密度関数$f(x)$、Yの周辺確率密度関数$g(y)$は、それぞれ次のように求めることができます。

$$f(x) = \int h(x, y) dy$$

$$g(y) = \int h(x, y) dx$$

- 離散型確率変数　－－－　連続型確率変数
  和のオペレータ　－－－　積分
  確率質量関数　－－－　確率密度関数

という感じで対応しています。以下では離散型のみについて議論します。

18

2つの確率変数の分布：これまでのまとめ

- X、Y：離散型確率変数
- 結合確率質量関数：$h(x, y)$
  - この関数のインプットとアウトプットは以下の通り。
  - インプット：xの具体的な数字、yの具体的な数字
  - アウトプット：$(X = x, Y = y)$の組み合わせが起こる確率
  - 結合確率分布を表す。
- Xの周辺確率質量関数：$f(x) = \sum_y h(x, y)$
  - この関数のインプットとアウトプットは以下の通り。
  - インプット：xの具体的な数字
  - アウトプット：$X = x$が起こる確率
  - Xの周辺確率分布を表す。Xの期待値は$E(X) = \sum_x x f(x)$。
- Yの周辺確率質量関数：$g(y) = \sum_x h(x, y)$
  - $Y = y$が起こる確率。Yの周辺確率分布を表す。Yの期待値は$E(Y) = \sum_y y g(y)$。

19

2つの確率変数：独立

- 離散型確率変数X、Y

これらの結合確率質量関数$h(x, y)$が、すべての実現値(x, y)において、それぞれの周辺確率質量関数$f(x)$と$g(y)$の積になっているとき、すなわち

$$h(x, y) = f(x) g(y)$$

のとき、XとYは互いに**独立**であると言います。

- 講義ノート「確率の基礎」で扱った「独立」の概念、思い出してください。

- 「XとYが独立」とは、XとYは全く何にも関係していない、というとても強い概念です。

20

確率変数の和･積の期待値

- X、Y: 離散型確率変数
- X、Yの結合確率質量関数：$h(x, y)$

このときXとYの和･積の期待値は、

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) h(x_i, y_j)$$
$$E(XY) = \sum_i \sum_j x_i y_j h(x_i, y_j)$$

- X+YやXYを一つの確率変数として考えれば、ここでやってることは一変数のときの期待値と本質的には同じ。

21

例：X＋Yの期待値

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) h(x_i, y_j)$$

$$\begin{aligned} E(X+Y) &= (1+5)\ h(1{,}5) + (1+10)\ h(1{,}10) + (2+5) h(2{,}5) + (2+10) h(2{,}10) \\ &= (1+5) \times\ 0.2\ \ + (1+10) \times 0.4\ \ \ + (2+5) \times 0.1\ \ + (2+10) \times 0.3 \\ &= \end{aligned}$$

22

例：XYの期待値

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) h(x_i, y_j)$$

$$\begin{aligned} E(XY) &= (1 \cdot 5)\ h(1{,}5) + (1 \cdot 10)\ h(1{,}10) + (2 \cdot 5) h(2{,}5) + (2 \cdot 10) h(2{,}10) \\ &= (1 \cdot 5) \times\ 0.2\ \ + (1 \cdot 10) \times 0.4\ \ \ + (2 \cdot 5) \times 0.1\ \ + (2 \cdot 10) \times 0.3 \\ &= \end{aligned}$$

23

確率変数の和の期待値

- 以下は回帰分析でとてもよく使う公式です。

X、Yがそれぞれ確率変数で、aとbが定数のとき
$$E(aX + bY) = aE(X) + bE(Y)$$

証明：$E(aX + bX)$

$$\begin{aligned} &= \sum_x \sum_y (ax + by) h(x, y) = a \sum_x \sum_y x h(x, y) + b \sum_x \sum_y y h(x, y) \\ &= a \sum_x x \boxed{\sum_y h(x, y)} + b \sum_y y \boxed{\sum_x h(x, y)} \\ &= a \boxed{\sum_x x f(x)} + b \boxed{\sum_y y g(y)} \\ &= aE(X) + bE(Y) \end{aligned}$$

この公式はXとYが独立か否かに関係なく、常に成り立ちます。

24

確率変数の和の期待値

- 例：Xをサイコロを三回投げたときの出た目の和とする。この期待値は？
- X_i (i = 1, 2, 3)をi番目のさいころの目とすると

$$X = X_1 + X_2 + X_3$$

この期待値は

$$E[X] = E[X_1] + E[X_2] + E[X_3]$$

サイコロを一回投げたときの出目の期待値は ….. だから

$$E[X_i] = \ldots$$

従って、

$$E[X] = \ldots$$

25

確率変数の積の期待値

- 積の期待値は少しトリッキーです。
- $E(XY) = E(X)E(Y)$となりそうなものですが、、、
- これは特殊なケースを除いては成り立ちません。
- $E(XY) = \sum_x \sum_y xyh(x,y)$
- $E(X)E(Y) = \left(\sum_x xf(x)\right)\left(\sum_y yg(x)\right)$
- これら二つは特殊なケースを除くと同じではありません。
- それではその特殊なケースとは？：独立なケースです。

26

X、Yが独立なら

$$E(XY) = E(X)E(Y)$$

X、Yが独立でないなら

$$E(XY) \neq E(X)E(Y)$$

- 証明：もしXとYが独立なら$h(x,y) = f(x)g(y)$だから、

$$E(XY) = \sum_x \sum_y xyh(x,y) = \sum_x \sum_y xyf(x)g(y)$$
$$= \sum_x xf(x) \sum_y yg(y) = \sum_x xf(x)E(Y) = E(Y) \sum_x xf(x) = E(Y)E(X)$$

27

確率変数の和と積の期待値：まとめ

- 和の期待値

$$E(X+Y) = E(X) + E(Y)$$

- 積の期待値：もしXとYが独立なら、

$$E(XY) = E(X)E(Y)$$

XとYが独立でないなら、

$$E(XY) \neq E(X)E(Y)$$

28

共分散

- 2つの確率変数の共変動を測る指標の一つが**共分散**。
- 確率変数XとYの共分散は以下のように定義される：

$$COV(X,Y) = E[(X - E(X))(Y - E(Y))]$$

- 共分散は以下のようにも表すことができる（証明せよ）：

$$COV(X,Y) = E(XY) - E(X)E(Y)$$

- もしXとYが独立なら$E(XY) = E(X)E(Y)$なので

$$COV(X,Y) = 0$$

- 注意：逆は必ずしも成り立たない。すなわち、$COV(X,Y) = 0$でも、XとYが独立とは限らない。

共分散

- 共分散が何を表すかを知るために次の例を考えましょう。

$$X = \begin{cases} 1 & if\ A\ occurs \\ 0 & otherwise \end{cases} \qquad Y = \begin{cases} 1 & if\ B\ occurs \\ 0 & otherwise \end{cases}$$

- 共分散は$COV(X,Y) = E(XY) - E(X)E(Y)$
- それぞれの項を計算してみます。
- $E(XY) = \sum_x \sum_y xyh(x,y) = 0 \cdot 0 \cdot h(0,0) + 1 \cdot 0 \cdot h(1,0) + 0 \cdot 1 \cdot h(0,1) + 1 \cdot 1 \cdot h(1,1)$
  $= h(1,1) = Pr(X = 1, Y = 1)$
- $E(X) = \sum_x xf(x) = 0 \cdot f(0) + 1 \cdot f(1) = f(1) = Pr(X = 1)$
- $E(Y) = \sum_y yg(y) = 0 \cdot g(0) + 1 \cdot g(1) = g(1) = Pr(Y = 1)$

共分散

- 従って

$$COV(X,Y) = Pr(X = 1, Y = 1) - Pr(X = 1)\,Pr(Y = 1)$$

- 「共分散が正である」とは何を意味する？

$$COV(X,Y) > 0 \to Pr(X = 1, Y = 1) > Pr(X = 1)\,Pr(Y = 1)$$

$Pr(X = 1, Y = 1) = Pr(Y = 1|X = 1)Pr(X = 1)$ だから（条件付確率を思い出してください）、

$$\to Pr(Y = 1|X = 1)Pr(X = 1) > Pr(X = 1)\,Pr(Y = 1)$$
$$\to Pr(Y = 1|X = 1) > Pr(Y = 1)$$

また$Pr(Y = 1) = Pr(Y = 1|X = 1)\,Pr(X = 1) + Pr(Y = 1|X = 0)\,Pr(X = 0)$なので、$Pr(Y = 1) > Pr(Y = 1|X = 0)$

合わせると、$Pr(Y = 1|X = 1) > Pr(Y = 1|X = 0)$

これは「Xが大きいとき（ここでは1）のときに、Yも大きくなる傾向にある」ということ。

- 一般的に、

$COV(X,Y) > 0$なら、「$X(Y)$が大きいときに、$Y(X)$も大きくなる傾向がある」ことを示す（**正の相関**があるという）。

$COV(X,Y) < 0$なら、「$X(Y)$が大きいときに、$Y(X)$は小さくなる傾向がある」ことを示す（**負の相関**があるという）。

確率変数の和の分散

- 二つの確率変数があるとして、、、、
- 和の期待値のルール：$E(X+Y)=E(X)+E(Y)$
- 積の期待値のルール：独立なら$E(XY)=E(X)E(Y)$。
  そうでなければ$E(XY)\neq E(X)E(Y)$
- 共分散（$COV(X,Y)=E[(X-E(X))(Y-E(Y))]$）$=E(XY)-E(X)E(Y)$
- これらは、回帰分析において非常に重要なルールやコンセプトです。
- もう一つ重要なものに、「確率変数の和の分散」があります。

$$VAR(X+Y)$$

33

確率変数の和の分散

思い出そう：確率変数の分散は
$VAR(X)=E[(X-E(X))^2]$

$VAR(X+Y)$
$=E\left[[(X+Y)-E(X+Y)]^2\right]$
$=E\left[[(X-E(X))+(Y-E(Y))]^2\right]$
$=E\left[(X-E(X))^2+2(X-E(X))(Y-E(Y))+(Y-E(Y))^2\right]$
$=E[(X-E(X))^2]+2E\left[(X-E(X))(Y-E(Y))\right]+E[(Y-E(Y))^2]$
$=VAR(X)+2COV(X,Y)+VAR(Y)$

となります。まとめると、

確率変数の和の分散の公式：

$$VAR(X+Y)=VAR(X)+2COV(X,Y)+VAR(Y)$$

確率変数が相関していないなら、$VAR(X+Y)=VAR(X)+VAR(Y)$

もちろん独立なら、$VAR(X+Y)=VAR(X)+VAR(Y)$

34

条件付き期待値

- 回帰分析の理論を学習するためのとりあえずの下準備の終了まであと少しです。
- 回帰分析で最も大切なコンセプトの1つが、**条件付き期待値**。
  - 独立変数がある値を取るとする。
  - このとき従属変数は平均するとどのような値をとる？
  - 「独立変数がある値を取る」という条件の下で、従属変数の期待値は？
  - これ、条件付き期待値。
- 回帰分析において、変数の関係のモデル化は、（多くの場合）条件付き期待値をどのようにモデル化するか、ということ。
- なので、条件付き期待値の概念を理解することは極めて重要。
- その前に、確率変数における条件付確率、やります。

35

条件付き確率質量関数

- Recall: 事象Fの条件の下で、事象Eの条件付き確率は

$$Pr(E|F)=\frac{Pr(E\cap F)}{Pr(F)}$$

- 2つの離散確率変数：X、Y
- 確率変数Xがxという値をとるという条件の下で、確率変数Yがyという値をとる確率は

$$Pr(Y=y|X=x)$$

- この確率を次のように表すとする。

$$Pr(Y=y|X=x)=g(y|x)$$

- $g(y|x)$を「Yの**条件付き確率質量関数**」と呼ぶ。

36

条件付き確率質量関数

- Recall: 事象Fの条件の下で、事象Eの条件付き確率は

$$Pr(E|F) = \frac{Pr(E \cap F)}{Pr(F)}$$

- $Pr(Y = y|X = x) = \frac{Pr(Y=y,X=x)}{Pr(X=x)}$

- よって、Yの条件付き確率質量関数$g(y|x)$は、XとYの結合確率質量関数$h(x, y)$とXの周辺確率質量関数$f(x)$と次のように関係している。

$$g(y|x) = \frac{h(x, y)}{f(x)}$$

37

条件付き確率質量関数:例

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

$$Pr(Y = 5|X = 1) = \frac{Pr(Y = 5, X = 1)}{Pr(X = 1)} = \frac{0.2}{0.6} = \frac{1}{3}$$

$$Pr(Y = 10|X = 1) = \frac{Pr(Y = 10, X = 1)}{Pr(X = 1)} = \frac{0.4}{0.6} = \frac{2}{3}$$

$$Pr(Y = 5|X = 2) = \frac{Pr(Y = 5, X = 2)}{Pr(X = 2)} = \frac{0.1}{0.4} = \frac{1}{4}$$

$$Pr(Y = 10|X = 2) = \frac{Pr(Y = 10, X = 2)}{Pr(X = 2)} = \frac{0.3}{0.4} = \frac{3}{4}$$

従って、Yの条件付確率質量関数$g(y|x)$は

$$\begin{aligned} g(y|x) &= \tfrac{1}{3},\ \ y = 5, x = 1 \\ &= \tfrac{2}{3},\ \ y = 10, x = 1 \\ &= \tfrac{1}{4},\ \ y = 5, x = 2 \\ &= \tfrac{3}{4},\ \ y = 10, x = 2 \end{aligned}$$

38

条件付き期待値

- $X = x$の下での、Yの**条件付き期待値**は次のように定義されます:

$$E(Y|X = x) = \sum_y y\, Pr(Y = y|X = x) = \sum_y y\, g(y|x)$$

- この定義は以前のYの期待値の定義と1つのことを除いて全く同じです。

- 違うところは、ここではすべてのことが「 $X = x$ 」という条件の下、という点です。

- この定義を使って、例における条件付き期待値を計算してみましょう。

39

条件付き期待値:例

$$E(Y|X = x) = \sum_y y\, Pr(Y = y|X = x) = \sum_y y\, g(y|x)$$

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

Yの条件付確率質量関数$g(y|x)$は

$$\begin{aligned} g(y|x) &= \tfrac{1}{3},\ \ y = 5, x = 1 \\ &= \tfrac{2}{3},\ \ y = 10, x = 1 \\ &= \tfrac{1}{4},\ \ y = 5, x = 2 \\ &= \tfrac{3}{4},\ \ y = 10, x = 2 \end{aligned}$$

$X = 1$という条件のもとでのYの期待値は?

$$\begin{aligned} E(Y|X = 1) &= 5 \cdot \Pr(Y = 5|X = 1) + 10 \cdot \Pr(Y = 10|X = 1) \\ &= 5 \cdot g(5|1) + 10 \cdot g(10|1) \\ &= 5 \cdot 1/3 + 10 \cdot 2/3 \\ &= 25/3 \end{aligned}$$

40

条件付き期待値:例

$$E(Y|X=x)=\sum_y y\,Pr(Y=y|X=x)=\sum_y y\,g(y|x)$$

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

Yの条件付確率質量関数$g(y|x)$は

$$g(y|x)=\frac{1}{3},\ \ y=5,x=1$$
$$=\frac{2}{3},\ \ y=10,x=1$$
$$=\frac{1}{4},\ \ y=5,x=2$$
$$=\frac{3}{4},\ \ y=10,x=2$$

$X=1$という条件のもとでのYの期待値は？

$$E(Y|X=1)=5\cdot\Pr(Y=5|X=1)+10\cdot\Pr(Y=10|X=1)$$
$$=5\cdot g(5|1)+10\cdot g(10|1)$$
$$=5\cdot 1/3+10\cdot 2/3$$
$$=25/3$$

$X=2$という条件のもとでのYの期待値は？

$$E(Y|X=2)=5\cdot\Pr(Y=5|X=2)+10\cdot\Pr(Y=10|X=2)$$
$$=5\cdot g(5|2)+10\cdot g(10|2)$$
$$=5\cdot 1/4+10\cdot 3/4$$
$$=35/4$$

41

条件付き期待値:例

$$E(Y|X=x)=\sum_y y\,Pr(Y=y|X=x)=\sum_y y\,g(y|x)$$

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

$X=1$という条件のもとでのYの期待値は？

$$E(Y|X=1)=5\cdot\Pr(Y=5|X=1)+10\cdot\Pr(Y=10|X=1)$$
$$=5\cdot g(5|1)+10\cdot g(10|1)$$
$$=5\cdot 1/3+10\cdot 2/3$$
$$=25/3$$

条件が変化すると、条件付き期待値の値も変化していることに注目。

$X=2$という条件のもとでのYの期待値は？

$$E(Y|X=2)=5\cdot\Pr(Y=5|X=2)+10\cdot\Pr(Y=10|X=2)$$
$$=5\cdot g(5|2)+10\cdot g(10|2)$$
$$=5\cdot 1/4+10\cdot 3/4$$
$$=35/4$$

42

条件付き期待値:例

$$E(Y|X=x)=\sum_y y\,Pr(Y=y|X=x)=\sum_y y\,g(y|x)$$

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

$X=1$という条件のもとでのYの期待値は？

$$E(Y|X=1)=5\cdot\Pr(Y=5|X=1)+10\cdot\Pr(Y$$
$$=5\cdot g(5|1)+10\cdot g(10|1)$$
$$=5\cdot 1/3+10\cdot 2/3$$
$$=25/3$$

ちなみにYの期待値は

$$E(Y)=5\cdot\Pr(Y=5)+10\cdot\Pr(Y=10)$$
$$=5\cdot g(5)+10\cdot g(10)$$
$$=5\cdot 0.3+10\cdot 0.7$$
$$=8.5$$

$X=2$という条件のもとでのYの期待値は？

$$E(Y|X=2)=5\cdot\Pr(Y=5|X=2)+10\cdot\Pr(Y=$$
$$=5\cdot g(5|2)+10\cdot g(10|2)$$
$$=5\cdot 1/4+10\cdot 3/4$$
$$=35/4$$

43

条件付き期待値: 独立のケース

$$E(Y|X=x)=\sum_y y\,Pr(Y=y|X=x)=\sum_y y\,g(y|x)$$

- 確率変数とX、Yが独立の場合、Yの条件付き期待値はどうなる？
- $g(y|x)=h(x,y)/f(x)$
- X、Yが独立の場合、$h(x,y)=f(x)g(y)$
- よってX、Yが独立なら、$g(y|x)=g(y)$
- これより$E(Y|X=x)=\sum_y y\,g(y|x)=\sum_y y\,g(y)=E(Y)$

確率変数とX、Yが独立の場合、Yの条件付き期待値は単なるYの期待値と同じになる。

44

条件付期待値の特性

- Yの条件付き期待値は、条件$(X = x)$が変わると、変化する。
- もう少し丁寧に表すと、$E(Y|X = x_1), E(Y|X = x_2), \ldots$の値は、$x_i(i = 1,2, \ldots)$の値が変わると、変化する。
  - 先ほどの例を見てみて下さい。
- 従って、$E(Y|X)$は、**<u>確率変数Xの関数</u>**と考えられる。
- 「$E(Y|X)$は **<u>確率変数Xの関数</u>**」、これを理解することは重要。
- 以下では、条件付き期待値に関する重要な特性を紹介していきます。
- 頻繁に使います。

条件付期待値の特性（繰り返し期待値の法則）

- 特性1: 繰り返し期待値の法則

$$E(Y) = E[E(Y|X)]$$

- 証明：Yが離散の時には右辺は

$$E[E(Y|X)] = \sum_x E(Y|X = x)Pr(X = x)$$

で、

$$\begin{aligned}\sum_x E(Y \mid X = x)P(X = x) &= \sum_x \sum_y yP(Y = y \mid X = x)P(X = x) = \sum_x \sum_y y \frac{P(Y = y, X = x)}{P(X = x)} P(X = x) \\ &= \sum_x \sum_y yP(Y = y, X = x) = \sum_y \sum_x yP(Y = y, X = x) = \sum_y y \sum_x P(Y = y, X = x) \\ &= \sum_y yP(Y = y) = E(Y)\end{aligned}$$

繰り返し期待値の法則：応用例

- $Y = WAGE$（時間当たり賃金＄）、$X = EDU$（教育年数）とする。
- EDUを条件としたときの$WAGE$の条件付き期待値が

$$E[WAGE \mid EDU] = 4 + 0.6EDU$$

で表されるとする。

- $E[EDU] = 11.5$（years）だとすると、$E[WAGE]$は？
- 繰り返し期待値の法則を使う：

$$\begin{aligned}E[WAGE] &= E[E[WAGE \mid EDU]] = E[4 + 0.6EDU] \\ &= 4 + 0.6E[EDU] = 4 + 0.6(11.5) = \$10.9\end{aligned}$$

条件付期待値の特性

- 特性2：いかなる関数$c(X)$について、

$$E(c(X)|X) = c(X)$$

- この特性の意味は、「Xを条件として期待値を計算する際には、Xの関数は定数のように扱ってよい」ということ。
- 例：　$E(X^2|X) = X^2$（Xを知っていればX^2は分かる。）

条件付期待値の特性

- 特性3：いかなる関数$a(X), b(X)$について、

$$E(a(X)Y+b(X)|X)=a(X)E(Y|X)+b(X)$$

- 例： $E(XY+X^2|X)=XE(Y|X)+X^2$

- 特性4：XとYが独立なら

$$E(Y|X)=E(Y)$$

- 特性4の特殊ケース：

UとXが独立で、かつ$E[U]=0$とする。このとき$E(U|X)=0$

条件付期待値の特性

- 特性5：もし$E[Y|X]=E[Y]$なら

$$COV(X,Y)=0$$

また、いかなるXの関数もYとは相関しない：

$$COV(c(X),Y)=0$$

- 証明：定義より $COV(X,Y)=E(XY)-E(X)E(Y)$ 。繰り返し期待値の法則より、

$$E(XY)=E\big(E(XY|X)\big)=E(XE(Y|X))$$

もし$E(Y|X)=E(Y)$なら、$E(XY)=E\big(XE(Y)\big)=E(X)E(Y)$。
従って$COV(X,Y)=0$。

- 特性5：「もし$E[Y|X]=E[Y]$なら$COV(X,Y)=0$」の<u>逆は必ずしも成り立たない。</u>

- すなわち、XとYが相関していないとしても、$E[Y|X]=E[Y]$は必ずしも成り立たない。$E[Y|X]$がXに依存することあり。

- 例：$Y=X^2$ とする。

- このとき$E(Y|X)=E(X^2|X)=X^2$

- この$E(Y|X)$は明らかにXの関数。だから$E(Y|X)=E(Y)$は成り立たない。

- しかし$COV(Y,X)=0$はあり得る（例えば$E(X)=E(X^3)=0$のとき）。

超重要

条件付き期待値の特性

- 特性１（繰り返し期待値の法則）と特性5は次の重要なインプリケーションを与える。

UとXが確率変数で、$E[U|X]=0$が成り立つとする。このとき

$$E(U)=0$$
$$COV(U,X)=0$$
$$E(XU)=0$$

が成り立つ。

証明：繰り返し期待値の法則より、$E(U)=E(E(U|X))=E(0)=0$。

$E(U|X)=0$ また$E(U)=0$なので、$E(U|X)=E(U)$。従って、特性5より、$COV(U,X)=0$。

$COV(X,U)=E(XU)-E(X)E(U)=0$。$E(U)=0$より$E(XU)=0$。

条件付き分散

- 「条件付」の最後のトピックになります、「条件付き分散」です。
- 確率変数X、Y
- $X = x$の条件の下での Yの条件付き分散は次のように定義されます。

$$VAR(Y|X = x) = E[(Y - E(Y|X = x))^2|X = x]$$

- 条件付き分散は次のようにも表すことができます:

$$VAR(Y|X = x) = E(Y^2|X = x) - [E(Y|X = x)]^2$$

- 見た目ややこしいですが、「条件付き」の部分以外は普通の分散と同じです。
- 条件付き分散の特性: XとYが独立なら、$VAR(Y|X) = VAR(X)$

53

条件付き分散:例

	Y = 5	Y = 10
X = 1	0.2	0.4
X = 2	0.1	0.3

Yの条件付確率質量関数$g(y|x)$は

$$g(y|x) = \tfrac{1}{3},\ \ y = 5, x = 1$$
$$= \tfrac{2}{3},\ \ y = 10, x = 1$$
$$= \tfrac{1}{4},\ \ y = 5, x = 2$$
$$= \tfrac{3}{4},\ \ y = 10, x = 2$$

$VAR(Y|X = 2) = E(Y^2|X = 2) - [E(Y|X = 2)]^2$

$E(Y|X = 2) = 35/4 \rightarrow [E(Y|X = 2)]^2 = 1225/16$

$$E(Y^2|X = 2) = 5^2 \cdot Pr(Y = 5|X = 2) + 10^2 \cdot Pr(Y = 10|X = 2)$$
$$= 5^2 \cdot g(5|2) + 10^2 \cdot g(10|2)$$
$$= 5^2 \cdot (1/4) + 10^2 \cdot (3/4) = 325/4$$

$VAR(Y|X = 2) = 325/4 - 1225/16 = 75/16$

54

まとめ

- 離散型確率変数X, Y
- Xと Yの結合確率質量関数: $Pr(X = x, Y = y) = h(x, y)$
- Yの周辺確率質量関数: $g(y) = \sum_x h(x, y)$
- Xの周辺確率質量関数: $f(x) = \sum_y h(x, y)$
- X, Yが独立 $\leftrightarrow$ $h(x, y) = f(x)g(y)$
- aとbが定数のとき $E(aX + bY) = aE(X) + bE(Y)$
- X, Yが独立なら$E(XY) = E(X)E(Y)$. 独立でないなら$E(XY) \neq E(X)E(Y)$.
- Xと Yの共分散: $COV(X, Y) = E[(X - E(X))(Y - E(Y))]$
  - $COV(X, Y) > 0$: Xと Yは正に相関する
  - $COV(X, Y) < 0$: Xと Yは負に相関する

55

- $VAR(X + Y) = VAR(X) + 2COV(X, Y) + VAR(Y)$
- X, Yが独立なら、$VAR(X + Y) = VAR(X) + VAR(Y)$
- Yの条件付確率質量関数: $Pr(Y = y|X = x) = \frac{h(x,y)}{f(x)} = g(y|x)$
- $X = x$の下での Yの条件付き期待値: $E(Y|X = x) = \sum_y y\, g(y|x)$
- X、Yが独立の場合、$E(Y|X) = E(Y)$
- 繰り返し期待値の法則: $E(Y) = E[E(Y|X)]$
- いかなる関数$c(X)$について、 $E(c(X)|X) = c(X)$
- いかなる関数$a(X), b(X)$について、$E(a(X)Y + b(X)|X) = a(X)E(Y|X) + b(X)$
- もし$E[Y|X] = E[Y]$なら$COV(X, Y) = 0$ （逆は必ずしも成り立たない）

56

- UとXが確率変数で、$E[U|X]=0$が成り立つとする。このとき$E(U)=0$、$COV(U,X)=0$、$E(XU)=0$が成り立つ。

- $X=x$の条件の下でのYの条件付き分散は、

$$VAR(Y|X=x)=E[(Y-E(Y|X=x))^2|X=x]$$

または

$$VAR(Y|X=x)=E(Y^2|X=x)-[E(Y|X=x)]^2$$

- XとYが独立なら、$VAR(Y|X)=VAR(X)$