23 mai 2013
Prob #06

確率論 (*Probability Theory*) 同時分布

w.yamamoto

1 参考書と授業内容の差異

累積分布関数 F

累積分布関数の定義は、次のように「確率変数 X が x 以下の値をとる確率」である。

$$F(x) = Pr[X \leq x] \tag{1}$$

X が離散確率変数の場合 (離散な確率分布の場合とも) には、一般には

$$Pr[X \leq x] \neq Pr[X > x] \tag{2}$$

$$Pr[X < x] = 1 - Pr[X \leq x] \neq 1 - Pr[X < x] = Pr[X \leq x] \tag{3}$$

である。X が連続確率変数の場合 (連続な確率分布の場合とも) には、

$$Pr[X \leq x] = Pr[X > x] \tag{4}$$

$$Pr[X < x] = 1 - Pr[X \leq x] = 1 - Pr[X < x] = Pr[X \leq x] \tag{5}$$

となることが多い。なぜか？

確率分布の定義 F

標本空間に自然な順序が定まっている場合、あらゆる確率は累積分布関数 $F(x)$ を与えれば一意に定まる。

下側確率 $Pr[X \leq b] = F(b)$

上側確率 $Pr[X > a] = F(a)$

区間確率 $Pr[a < X \leq b] = F(b) - F(a)$

離散確率変数については、例えば整数値をとる場合は、x が整数ならば

$$Pr[X = x] = Pr[x - 1/2 \leq X \leq x + 1/2] = Pr[x - 1 < X < x + 1] \tag{6}$$

である。

原点モーメント m_k

$$m_k E\left[X^k\right] = \begin{cases} \sum_{x=-\infty}^{\infty} x^k p(x) & (X\text{ が離散確率変数の場合}) \\ \int_{x=-\infty}^{\infty} x^k f(x)\,dx & (X\text{ が連続確率変数の場合}) \end{cases} \tag{7}$$

1. m_1 は「原点との差」の確率分布の中心 ($p(x)$ や $f(x)$ をウェイトとして求めた、標本空間の重心)。
2. m_2 は「原点からの 2 乗距離」の確率分布の中心 ($p(x)$ や $f(x)$ をウェイトとして求めた、標本空間の原点からの 2 乗距離、の重心)。原点からの 2 乗距離は、確率分布の幅を広げても大きくなるし、確率分布の中心 m_1 を原点から離しても大きくなる。
3. $m_k, k \geq 3$ はそれぞれ「原点からの 3 乗距離」、「原点からの 4 乗距離」、... の確率分布の中心。

上で出て来た「の確率分布の中心」という言葉は、確率論では「期待値」と短く呼ばれる。以下では、期待値、という言葉の方を用いていく。また一般に、確率分布 F の期待値を μ、分散を σ^2 と書く。μ は上の m_1 のことであり、後述する $\mu_1, \mu_2, \ldots$ とは異なるので注意すること。

期待値 $E[\cdot]$

何かの確率分布の中心 (離散の値をとる何かであれば確率関数を、連続の値をとる何かであれば確率密度関数を、それぞれウェイトとして求めた重心のこと) を計算したものを、期待値という。確率分布 F の m_1 のことを特に、「F の期待値」と呼び、μ と記す。確率変数 X に関して期待値を計算することを $E_X[\cdot]$、確率分布 F に照らして期待値を計算することを $E_F[\cdot]$ と記す。2 つの確率変数があり、それぞれが $X \sim F$、$Y \sim G$ という確率分布に従っているとき、X と Y が互いに全く影響を及ぼし合わないならば、

$$E_X[X] = \mu_X \tag{8}$$

$$E_X[Y] = Y \tag{9}$$

$$E_Y[X] = X \tag{10}$$

$$E_Y[Y] = \mu_Y \tag{11}$$

である。

中心モーメント m_k

$$\mu_k E\left[(X - m_1)^k\right] = \begin{cases} \sum_{x=-\infty}^{\infty} (x-\mu)^k p(x) & (X\text{が離散確率変数の場合}) \\ \int_{x=-\infty}^{\infty} (x-\mu)^k f(x)\,dx & (X\text{が連続確率変数の場合}) \end{cases} \tag{12}$$

1. μ_1 は「$m_1 = \mu$ との差」の期待値 ($p(x)$ や $f(x)$ をウェイトとして求めた、標本空間の重心を $m_1 = \mu$ にずらした後での重心) であり、$E[X - \mu_1] = mu_1 - \mu_1$ より常に 0。

2. μ_2 は「$m_1 = \mu$ からの 2 乗距離」の確率分布の中心 ($p(x)$ や $f(x)$ をウェイトとして求めた、確率分布の中心 μ からの 2 乗距離、の重心)。これは確率分布の中心 $m_1 = \mu$ を原点から離しても変化せず、確率分布の幅を広げたときだけ大きくなる。

3. $\mu_k, k \geq 3$ はそれぞれ「$m_1 = \mu$ からの 3 乗距離」、「$m_1 = \mu$ からの 4 乗距離」、... の確率分布の中心で、3 次のモーメント、4 次のモーメントなどと呼ばれる。

歪度 β_1

確率分布が中心に対して対象かどうかを見る指標。定義は

$$\beta_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{{\mu_2}^{3/2}} \tag{13}$$

尖度 β_2

確率分布が尖り具合を見る指標。定義は次の 2 つがよく用いられている。

$$\beta_1 = \begin{cases} E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{{\mu_2}^2} \\ E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mu_4}{{\mu_2}^2} - 3 \end{cases} \tag{14}$$

両者の違いは $\mu_4/{\mu_2}^2$ の値を 3 と比較するか、 $\mu_4/{\mu_2}^2 - 3$ の値を 0 と比較するか、のみ。

位置変換と尺度変換

確率論では頻繁に、確率分布の中心を動かしたり、確率分布の広がり具合を変えたりする。この 2 つの操作にはそれぞれ、位置変換と尺度変換という名前がついている。位置変換は $y_1 = x + b$ のように、元の変数に b を加えて中心をずらす変換であり、$x = y_1 - b$ を代入するから

$$F(x) \Rightarrow F(y_1 - b) \tag{15}$$

となる。尺度変換は $y_2 = ax$ のように、元の変数に a を掛けて原点からの距離を遠くする変換であり、$x = y_2/a$ を代入するから

$$F(x) \Rightarrow F(y_2/a) \tag{16}$$

となる。

　位置変換をしてから尺度変換をした結果と、尺度変換をしてから位置変換をした結果は、同じになるか？答えは、ならない、が正しい。それぞれの変換は

$$Y_{12} = a\left(X + b\right) \tag{17}$$

$$Y_{21} = aX + b \tag{18}$$

であり、それぞれの結果は

$$F_{12}\left(y_{12}\right) = F\left(\frac{y_{12}}{a} - b\right) \tag{19}$$

$$F_{21}\left(y_{21}\right) = F\left(\frac{y_{21} - b}{a}\right) \tag{20}$$

となる。

確率論　第６週　同時分布

学籍番号 ____________________ 氏名 ____________________

6. 同時分布の取り扱いの練習

確率ベクトル (X, Y) の同時確率密度関数として、次の 2 つを考える。

$$f_{X,Y}(x, y) = \begin{cases} x + y & (x, y) \in D \text{ のとき} \\ 0 & (x, y) \notin D \text{ のとき} \end{cases}$$

$$g_{X,Y}(x, y) = \begin{cases} xy + \frac{1}{2}(x + y) + \frac{1}{4} & (x, y) \in D \text{ のとき} \\ 0 & (x, y) \notin D \text{ のとき} \end{cases}$$

ここで、$D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ である。このとき、次の問いに答えよ。

5-1 密度関数が $f_{X,Y}$ となる (X, Y) の同時確率分布 F について

5-1-a X のみの周辺分布の周辺密度関数 f_X を求めよ。

Ans. $f_X(x) =$ ____________________

5-1-b 同じく Y の周辺分布の密度関数 f_Y を求めよ。

Ans. $f_Y(y) =$ ____________________

5-1-c X と Y が互いに独立か調べよ。

Ans. ____________________

5-1-d $Pr[X + Y < 1]$ を求めよ。

Ans. $Pr[X + Y < 1] =$ ____________________

5-2 密度関数が $g_{X,Y}$ となる (X, Y) の同時確率分布 G について

5-2-a X の周辺分布の密度関数 g_X を求めよ。

Ans. $g_X(x) =$ ____________________

5-2-b Y の周辺分布の密度関数 g_Y を求めよ。

Ans. $g_Y(y) =$ ____________________

5-2-c X と Y が互いに独立か調べよ。

Ans. ____________________

5-2-d $Pr[X + Y < 1]$ を求めよ。

Ans. $Pr[X + Y < 1] =$ ____________________

連絡 学籍番号が 12 以外で始まる学生が、4 限の確率論の講義の履修を希望する場合、今年度は 5 限に上級科目を履修する場合のみ許可する。それ以外の学生は必ず 5 限を履修すること。(なお参考までに次年度以降はいかなる理由でも許可しない。)

レポートの提出要領は次の通り。

課題番号	#6 (2013.05.23 出題)
提出期限	2013 年 5 月 27 日 午後 4 時 30 分
提出場所	西 5 号館 3 階総合情報学科事務室の向かい側の集合ポスト (「確率論」あるいは「応用数学Ｂ」とある投函口)
様　　式	本紙、Ａ４もしくはＢ５（ルーズリーフ可、両面可）
そ の 他	丸写しは採点していて飽きるし、剽窃は自分のためにならない 各自が自力で取り組むことを、切に願う 成書を参考にするなとは言わないが、参考にした書籍があれば、著者への礼儀として必ず記すこと 表紙はつけないこと １ページ目の上部に、「講義名」「レポート番号」「学籍番号」「氏名」「投函日」を記すこと

確率論　第６週　同時分布

学籍番号 ＿＿＿＿＿＿＿＿＿＿ 氏名 ＿＿＿＿＿＿＿＿＿＿

6. 同時分布の取り扱いの練習

確率ベクトル (X, Y) の同時確率密度関数として、次の 2 つを考える。

$$f_{X,Y}(x, y) = \begin{cases} x + y & (x, y) \in D \text{ のとき} \\ 0 & (x, y) \notin D \text{ のとき} \end{cases}$$

$$g_{X,Y}(x, y) = \begin{cases} xy + \frac{1}{2}(x + y) + \frac{1}{4} & (x, y) \in D \text{ のとき} \\ 0 & (x, y) \notin D \text{ のとき} \end{cases}$$

ここで、$D = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1\}$ である。このとき、次の問いに答えよ。

5-1 密度関数が $f_{X,Y}$ となる (X, Y) の同時確率分布 F について

5-1-a X のみの周辺分布の周辺密度関数 f_X を求めよ。

Ans. $f_X(x) =$ ＿＿＿＿＿＿＿＿＿＿

5-1-b 同じく Y の周辺分布の密度関数 f_Y を求めよ。

Ans. $f_Y(y) =$ ＿＿＿＿＿＿＿＿＿＿

5-1-c X と Y が互いに独立か調べよ。

Ans. ＿＿＿＿＿＿＿＿＿＿

5-1-d $Pr[X + Y < 1]$ を求めよ。

Ans. $Pr[X + Y < 1] =$ ＿＿＿＿＿＿＿＿＿＿

5-2 密度関数が $g_{X,Y}$ となる (X, Y) の同時確率分布 G について

5-2-a X の周辺分布の密度関数 g_X を求めよ。

Ans. $g_X(x) =$ ＿＿＿＿＿＿＿＿＿＿

5-2-b Y の周辺分布の密度関数 g_Y を求めよ。

Ans. $g_Y(y) =$ ＿＿＿＿＿＿＿＿＿＿

5-2-c X と Y が互いに独立か調べよ。

Ans. ＿＿＿＿＿＿＿＿＿＿

5-2-d $Pr[X + Y < 1]$ を求めよ。

Ans. $Pr[X + Y < 1] =$ ＿＿＿＿＿＿＿＿＿＿

連絡 学籍番号が 12 以外で始まる学生が、4 限の確率論の講義の履修を希望する場合、今年度は 5 限に上級科目を履修する場合のみ許可する。それ以外の学生は必ず 5 限を履修すること。(なお参考までに次年度以降はいかなる理由でも許可しない。)

レポートの提出要領は次の通り。

課題番号	#6 (2013.05.23 出題)
提出期限	2013 年 5 月 27 日 午後 4 時 30 分
提出場所	西 5 号館 3 階総合情報学科事務室の向かい側の集合ポスト (「確率論」あるいは「応用数学Ｂ」とある投函口)
様　　式	本紙、Ａ４もしくはＢ５（ルーズリーフ可、両面可）
そ の 他	丸写しは採点していて飽きるし、剽窃は自分のためにならない 各自が自力で取り組むことを、切に願う 成書を参考にするなとは言わないが、参考にした書籍があれば、著者への礼儀として必ず記すこと 表紙はつけないこと １ページ目の上部に、「講義名」「レポート番号」「学籍番号」「氏名」「投函日」を記すこと