統計学（第二章　確率論、検定）

劉慶豊[1]

小樽商科大学

June 8, 2012

[1]E-mail:qliu@res.otaru-uc.ac.jp, URL:http://www.otaru-uc.ac.jp/~qliu/

2 変数同時分布表

例 15カ国の出生率と就学率の同時分布表

出生率＼就学率	40−50	50-60	60-70	70-100	出生率の周辺度数
10-20	0	1	2	3	6
20-30	3	2	1	2	8
30-40	1	0	0	0	1
就学率の周辺度数	4	3	3	5	15

出生率が $10-20\%$ で就学率が 70% 以上の国は3カ国であることが読み取れる。

周辺度数 各変数の周辺度数は1変数の度数分布表の時の度数と同じ。

確率入門

試行 結果に不確実性が伴うような実験や行動などを試行という。物理実験、政策施行、社会現象と自然現象などの発生を試行とみなすことができる。

事象 試行の結果を事象という。

根源事象 分割不可能な事象。

空事象 空である空集合。ϕで表す。

標本空間 すべての根元事象により構成した集合。標本空間は通常Ωで表す。

全事象 標本空間の中のすべての根元事象を含む事象である。

サイコロの例 サイコロを1回だけ投げる実験を行う。この実験が試行という。根元事象は,1の目が出る、2の目が出る、$\cdots$で、数学記号で表現すれば$A_1 = 1, A_2 = 2, \cdots, A_6 = 6$となる。標本空間は$\Omega = \{1, 2, 3, 4, 5, 6\}$で、$E_1 = \{1, 3, 5\}$という事象は奇数の目が出る結果を意味する。

確率

主観的確率 主観的に考えた事件の発生する可能性。例、明日の試合で勝利する確率は80%だ。彼が100%遅刻するだろう。

統計学的な確率の定義 同じ条件のもとで、繰り返し同じ試行をn回行い、毎回の試行は互いにまったく無関係（独立）であるとする。事象A（Aという結果）がm回出たとする。Aが発生する割合はm/nである。もし、試行の回数nを無限回にしたとする、このとき、m/nが一つの定数に収束したら、その定数を事象Aの確率とする。通常$P(A)$で表す。

練習問題 サイコロの例で統計学的な確率の概念を考えてみる。

確率の公理

確率の定義の仕方が様々あるが、統計学で取り扱えるようにするためには、必ず以下の公理を満たさなければならない。

確率の公理

1. 任意の事象Aに関して、$0 \leq P(A) \leq 1$。
2. 全事象の確率が1である。$P(\Omega) = 1$。
3. 事象AとBが排反する場合（同じ根元事象を含まない、すなわち、AとBが必ず同時に起きない場合）、AとBがどちらかが起きる確率はAとBの確率の和である。$P(A \cup B) = P(A) + P(B)$。$A$と$B$が排反ではない場合$P(A \cup B) = P(A) + P(B) - P(A \cap B)$である。

根元事象が互いに排反である。

確率変数

- 確率変数が取る値はおのおのに一定な確率が対応している。
- コイン投げの実験では表を1とし裏を0としたら、コイン投げの結果を確率変数Xのとる値とそれに対応している確率は$X = 0$の確率は$0.5, X = 1$の確率も0.5である。
- サイコロの場合は$X = 1$の確率は$1/6, X = 2$の確率も$1/6, \ldots$。

実現値　試行が完了して、確率変数の値が特定の実数値に確定される。その実数値が実現値という。

離散確率変数

- 上述した二つの確率変数の例では確率変数Xは0と1または1, 2, 3, 4, 5, 6と飛び飛びとした値しかとらない、たとえば0と1の間の値を取ることがない、この場合は離散確率変数と呼ぶ。
- 確率変数が取る値とそれに対応する確率との関係は以下の表で表せる。

Xが取る値	x_1	x_2	$\cdots$	x_n
対応する確率$P(x)$	$P(x_1)$	$P(x_2)$	$\cdots$	$P(x_n)$

サイコロの例

Xが取る値	$x_1=1$	$x_2=2$	$\cdots$	$x_6=6$
対応する確率$P(x)$	$P(x_1)=1/6$	$P(x_2)=1/6$	$\cdots$	$P(x_6)=1/6$

等確率の世界での離散確率変数の確率の計算

等確率の世界 すべての根元事象の確率が同じである世界。

例 歪みのないコイン、正確の形をしたサイコロ。

確率の計算 根元事象の確率が1/根元事象の総数で、個々の根元事象が互いに排反であるため

$$\text{事象}A\text{の確率} = \frac{\text{事象}A\text{に含まれた根元事象の数}}{\text{根元事象の総数}}$$

練習問題 2回コインを投げて、$A=\{$1回目に表が出て、2回目に裏が出る$\}$の確率を計算してください。

樹形図による確率の計算

樹形図 簡単な問題のとき、樹形図によって確率を計算できる。

根元事象の総数場合の数の計算と同じように、樹形図でできる。

しかし、場合の数が多すぎるとき困難である。

- 例：袋の中に白玉が1個、青玉が2個があり、玉2個を取り出すときに、全部青玉になる確率$=1/3$。テキスト図2.4参照。

順列

順列（順序が関係する場合に利用する） n個の異なった積み木があるとする。その中からr個を取り出して1列に並べたものを順列と呼ぶ。異なった順列の数は以下の公式で計算できる。

順列の数

$$_nP_r = n \times (n-1) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

ただし$n! = 1 \times 2 \times \cdots \times n$で$n$の階乗と呼ぶ。$0! = 1$とする。

例 A, B, C, Dの4人がいる場合、3人の列を作って、何パターンができるか？$_4P_3 = 4 \times 3 \times 2 = 24$。各自で樹形図で確認しなさい。

組合せ

組合せ（順序が関係しない場合に利用する） n個の異なった積み木があるとする。その中から r個を取り出して順番を無視してその r個の内容を組合せという。組合せの数は以下の公式で計算する。

組合せの数

$$_{n}C_{r} = \frac{n!}{r! \times (n-r)!} = \frac{_{n}P_{r}}{r!}$$

例 A, B, C, Dの 4 人がいる場合、3人を選ぶ場合、何パターンがあり得るのか？$_{4}C_{3} = (4 \times 3 \times 2) / (3 \times 2 \times 1) = 4$。各自で樹形図で確認しなさい。

順列、組合せによる確率の計算

例 袋の中に白玉が 1個、青玉が 2個があり、玉2個を取り出すときに、全部青玉になる確率を計算しなさい。

確率の公式事象Aの確率 $= \frac{\text{事象}A\text{に含まれた根元事象の数}}{\text{根元事象の総数}}$ と組合せの公式を利用して計算する

$$\begin{aligned}\text{全部青玉になる確率} &= \frac{\text{全部青玉になるパターンの数}}{\text{すべての可能なパターンの数}} \\ &= \frac{1}{_{2}C_{3}} = \frac{1}{3}\end{aligned}$$

例 サイコロを 2回投げて、目の合計が 10になる確率。順序を配慮に入れて 10になるパターンの数は 3、すべてのパターンの数は $_{1}C_{6} \times {}_{1}C_{6} = 36$、求める確率は $1/12$。

順列、組合せによる確率の計算

例（テキスト例２．７） 壺に4個の青玉と2個の赤玉があり、6個の玉を順番に取り出して並べ、2個の赤玉が続く確率は？

1. 2個の赤玉が青玉との相対的な位置を考えると5ヶ所しかない。
2. そして赤玉の順番が${}_2P_2 = 2$パターン。
3. さらに残りの4個の青玉の順番のパターンは${}_4P_4 = 4!$。
4. 合わせて2個の赤玉続けて出てくるパターンは $5 \times 2 \times 4!$。
5. 全ての可能性は${}_6P_6 = 6!$。
6. ゆえに、求める確率は$(5 \times 2 \times 4!)/6! = 1/3$。

条件付確率

条件付確率 事象Aが生じたという条件のもとで事象Bが生じる確率をAのもとでのBの条件付確率という。$P(B|A)$と記する。$P(A) \neq 0$の場合

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

と定義する。

例 シートベルト着用率を50%とし、シートベルトを着用し事故死の確率を0.5%とする。シートベルトを着用した人でその人の事故死の確率は？

$$\begin{aligned} & P(\text{事故死}|\text{シートベルト着用}) \\ &= \frac{P(\text{事故死} \cap \text{シートベルト着用})}{P(\text{シートベルト着用})} = \frac{0.5\%}{50\%} = 1\% \end{aligned}$$

ベイズの定理

ベイズの定理

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|A^C)\,P(A^C)}$$

ただし、A^C は A の補集合を現す。条件付確率の定義式と $P(B) = P\left[(B \cap A) \cup \left(B \cap A^C\right)\right] = P(B \cap A) + P\left(B \cap A^C\right)$ を利用して証明せよ。テキスト73ページ参照。

独立

独立 事象 A と B に関して、$P(A \cap B) = P(A) \times P(B)$ のとき、A と B が互いに独立であるという。独立であれば $P(B|A) = P(B)$。

例 コインを2回投げる事件を行う。事象 $A = \{$ 1回目には表が出る $\}$、事象 $B = \{$2回目には表が出る$\}$、特別なことをしない限り、2回の投げは互いに無関係で、独立である。$P(A \cap B) = P(A) \times P(B) = 0.5 \times 0.5 = 0.25$。

度数関数、分布関数

確率度数関数 ここでは $P(x)$ は確率度数関数と呼ばれる。$P(x_1), P(x_2)\ldots$ は x の具体的な値に対応する確率を表す。

分布関数 分布関数の定義は $F_X(c) = P(\{X \le c\})$ となる。表しているのは X が c より小さくなる確率である。離散確率変数の場合は

$$F_X(c) = P(\{X \le c\}) = \sum_{x_i \le c} P(x_i)$$

サイコロの例

$$\begin{aligned} F_X(3) &= P(\{X \le 3\}) = \sum_{x_i \le 3} P(x_i) \\ &= P(x_1) + P(x_2) + P(x_3) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \end{aligned}$$

確率変数の独立

確率変数の独立 任意の実数 a と b に関して、$P(\{X > a \text{かつ} Y > b\}) = P(\{X > a\}) \times P(\{Y > b\})$ が満たされるとき、X と Y が互いに独立であるという。直感的にいうと、X と Y が独立であるときは、X の実現値の大小は Y の実現値の大小とまったく無関係である。

確率変数の期待値

確率変数Xの期待値を $E(X)$で表す。期待値は直感的に確率変数の理論上の平均で理解できる。離散確率変数の場合

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

サイコロの例

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} x_i P(x_i) \\ &= P(x_1) \times x_1 + P(x_2) \times x_2 + \cdots + P(x_6) \times x_6 \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} \cdots + 6 \times \frac{1}{6} = \frac{7}{2} \end{aligned}$$

期待値の性質

- 以下の性質は後述する連続確率変数に関しても成り立つ。

期待値の加法性 任意の二つの確率変数Xと Yの和の期待値が期待値の和と等しい。$E(X+Y) = E(X) + E(Y)$。

独立な確率変数の積の期待値 確率変数Xと Yが独立であるとき $E(X \times Y) = E(X) \times E(Y)$。

定数倍 $E(cX) = cE(X)$、ただし cが定数である。

練習問題 以上の性質を利用して、$E(X+Y+Z+W) = E(X) + E(Y) + E(Z) + E(W)$を証明しなさい。$X, Y, Z$が互いに独立であるとする、$E(XYZ) = E(X)E(Y)E(Z)$を証明しなさい。

確率変数の分散

確率変数Xの分散を$V(X)$で表して、離散確率変数の場合

$$V(X) = E\left[(X - E(X))^2\right] = \sum_{i=1}^{n}\left[(x_i - E(X))^2 P(x_i)\right].$$

性質 $V(cX+b) = c^2 V(X)$、ただしcとbは定数である。連続確率変数に関しても成り立つ。

サイコロの例

$$\begin{aligned} V(X) &= \sum_{i=1}^{n}\left[(x_i - E(X))^2 P(x_i)\right] \\ &= (x_1 - E(X))^2 \times P(x_1) + \cdots + (x_6 - E(X))^2 \times P(x_6) \\ &= \left(1 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \cdots + \left(6 - \frac{7}{2}\right)^2 \times \frac{1}{6} \end{aligned}$$

確率変数の共分散

確率変数XとYの共分散を$Cov(X,Y)$で表して、離散確率変数の場合

$$\begin{aligned} Cov(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}\left[(x_i - E(X))(y_j - E(Y)) P(x_i, y_j)\right]. \end{aligned}$$

独立な確率変数の共分散 確率変数XとYが独立であるとき共分散$Cov(X,Y) = 0$。

確率変数の相関係数

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

$$-1 \le \rho(X,Y) \le 1$$

確率分布

Definition (確率分布)

ある確率変数Xの取る値に対応する確率がある関数$P(x)$で表せるとき、この確率変数Xの確率分布は$P(x)$に従うという。

ベルヌーイ分布の例

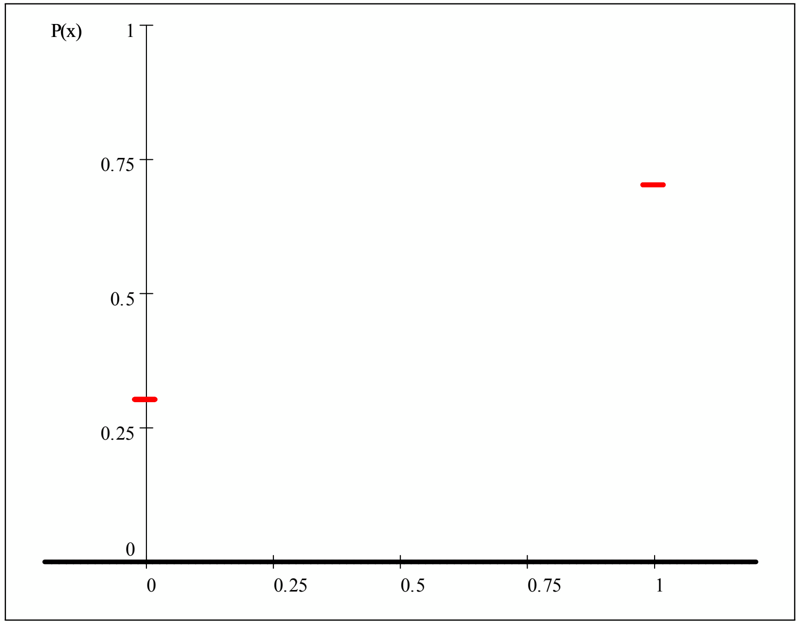
- ベルヌーイ分布の確率度数関数

$$\begin{cases} P(x) = p & x = 1 \\ P(x) = 1 - p & x = 0 \end{cases}$$

  $p = 0.5$のときは1回コイン投げた時の結果の分布となる。
- コインを投げる実験は確率$p = 0.5$のベルヌーイ試行と呼ぶ。
- 期待値$E(X) = 1 \times p + 0 \times (1 - p) = p$
- 分散は

$$\begin{aligned} V(X) &= (0 - p)^2 \times (1 - p) + (1 - p)^2 \times p \\ &= p^2 - p^3 + p - 2p^2 + p^3 = p - p^2 \end{aligned}$$

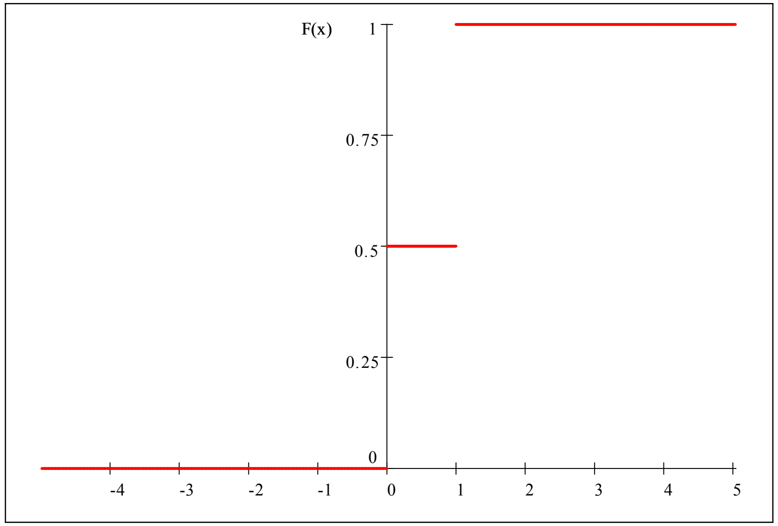
- 確率度数関数のグラフ



ベルヌーイの確率度数関数　$p = 0.7$

- 確率分布関数

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$



ベルヌーイ分布の分布関数 $p = 0.5$

二項分布

二項分布 独立な成功確率が p のベルヌーイ分布を n 回行って、成功した回数の分布が二項分布と呼ぶ。$B(n, p)$ で表記する。

二項分布の確率関数

$$P(x) = {}_nC_x p^x q^{n-x}, \ x = 0, 1, \cdots, n.$$

例 歪みのないコインを10回投げて、表が出る回数を X と記する場合、X の分布は $p = 0.5$ の二項分布となる。8回表が出る確率

$$P(X = 8) = {}_{10}C_8 p^8 (1 - p)^{10-8} = {}_{10}C_8 0.5^{10}$$

例 成功率 $p = 70\%$ バスケット選手の100回シュートして、成功する回数を X とする、X の分布は $p = 0.7$ の二項分布となる。この選手は100回シュートして80回成功する確率は

$$P(X = 80) = {}_{100}C_{80} p^{80} (1 - p)^{100-80} = {}_{100}C_{80} 0.7^{80} 0.3^{20}$$

二項分布の再生性 2つの二項分布の和がまた二項分布になる。

ポアソン分布

ポアソン分布 所与の時間間隔やある空間範囲内などでおきる事象の発生する確率を分析するためによく用いられる分布の一種。

ポアソン分布の確率関数

$$p(x) = \frac{\lambda^x \exp(-\lambda)}{x!}$$

例 過去の大阪市の交通事故の平均死亡者数は一日当たり 1 人だった。週間当たりの死亡者数を X とする。週間当たりの平均は 7 となるので、X は $\lambda = 7$ のポアソン分布に従う。一週間に 5 人以内になる確率は

$$\sum_{x=0}^{5} p(x) = \frac{7^0 \exp(-7)}{0!} + \frac{7^1 \exp(-7)}{1!} + \cdots + \frac{7^5 \exp(-7)}{5!}$$