二項分布・大数の法則

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I L07(2017-11-08 Wed)

最終更新: Time-stamp: "2017-11-08 Wed 07:15 JST hig"

今日の目標

- 二項分布の母期待値が計算できる 西川確率統計 §2.2.1
- チェビシェフの不等式から母平均値・母分散の意味を説明できる 西川確率統計 §2.3
- 大数の法則の意味が説明できる 西川確率統計 §4.1

http://hig3.net

L06-Q1

Quiz 解答:多次元の確率変数の期待値

1. $\mathrm{E}[X+2Y] = 0\cdot(1+2\cdot 0) + \frac{2}{12}(2+2\cdot 0) + \frac{1}{12}(3+2\cdot 0) + \frac{4}{12}(1+2\cdot 2) + 0(2+2\cdot 2) + \frac{5}{12}(3+2\cdot 2) = \frac{62}{12}$.
2. $\mathrm{E}[\mathbf{1}_{[Y\geq 1]}(X,Y)] = 0\cdot 0 + \frac{2}{12}\cdot 0 + \frac{1}{12}\cdot 0 + \frac{4}{12}\cdot 1 + 0\cdot 1 + \frac{5}{12}\cdot 1 = \frac{9}{12}$.
3. $$f_{\mathrm{X}}(x) = \begin{cases} 4/12 & (x=1) \\ 2/12 & (x=2) \\ 6/12 & (x=3) \\ 0 & (\text{他}) \end{cases} \qquad f_{\mathrm{Y}}(y) = \begin{cases} 3/12 & (y=0) \\ 9/12 & (y=2) \\ 0 & (\text{他}) \end{cases}$$
4. (1 の別解)

L06-Q4
Quiz 解答:多次元の確率変数の期待値

1. $\mathrm{E}[X+2Y] = 0\cdot(1+2\cdot 0) + \frac{2}{12}(2+2\cdot 0) + \frac{1}{12}(3+2\cdot 0) + \frac{4}{12}(1+2\cdot 2) + 0(2+2\cdot 2) + \frac{5}{12}(3+2\cdot 2) = \frac{62}{12}$.
2. $\mathrm{E}[\mathbf{1}_{[Y\geq 1]}(X,Y)] = 0\cdot 0 + \frac{2}{12}\cdot 0 + \frac{1}{12}\cdot 0 + \frac{4}{12}\cdot 1 + 0\cdot 1 + \frac{5}{12}\cdot 1 = \frac{9}{12}$.
3. $$f_{\mathrm{X}}(x) = \begin{cases} 4/12 & (x=1) \\ 2/12 & (x=2) \\ 6/12 & (x=3) \\ 0 & (\text{他}) \end{cases} \qquad f_{\mathrm{Y}}(y) = \begin{cases} 3/12 & (y=0) \\ 9/12 & (y=2) \\ 0 & (\text{他}) \end{cases}$$
4. (1 の別解)

L06-Q6
Quiz 解答:独立と限らない確率変数の母期待値

1. $\mathrm{E}[-2X+3Y] = -2\mathrm{E}[X] + 3\mathrm{E}[Y] = 5.$
2. $\mathrm{V}[-2X+3Y] = \mathrm{E}[(-2X+3Y)^2] - \mathrm{E}[-2X+3Y]^2 = (-2)^2\mathrm{V}[X] + 2(-2)(3)\mathrm{Cov}[X,Y] + 3^2\mathrm{V}[X] = 20 - 84 + 99 = 35.$

L06-Q7

Quiz 解答:2 つの離散型確率変数の母期待値・母平均値・母共分散・確率・独立性

$\mathrm{E}[X] = \frac{1}{7}\cdot 1 + \frac{6}{7}\cdot 3 = \frac{19}{7},$
$\mathrm{E}[Y] = \frac{3}{7}\cdot 2 + \frac{4}{7}\cdot 4 = \frac{22}{7},$
$\mathrm{E}[X^2] = \frac{1}{7}\cdot 1^2 + \frac{6}{7}\cdot 3^2 = \frac{55}{7},$
$\mathrm{E}[XY] = \frac{1}{7}\cdot 2\cdot 1 + \frac{2}{7}\cdot 3\cdot 2 + \frac{4}{7}\cdot 3\cdot 4 = \frac{62}{7}.$

1. 独立でない.
2. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{24}{49}.$
3. $\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = \frac{16}{49}.$

L06-Q8

Quiz 解答:離散型確率変数の独立性

確率の和は 1 なので, $\frac{2}{12} + \frac{1}{12} + A + B = 1$.
よって,

$$f_{\mathrm{Y}}(y) = \begin{cases} \frac{3}{12} & (y = 3) \\ \frac{9}{12} & (y = 7) \end{cases}$$

独立性から,

$$\begin{aligned} f_{\mathrm{XY}}(2,3) =& f_{\mathrm{X}}(2)\frac{3}{12} = \frac{2}{12}, \\ f_{\mathrm{XY}}(3,3) =& f_{\mathrm{X}}(3)\frac{3}{12} = \frac{1}{12}, \\ f_{\mathrm{XY}}(2,7) =& f_{\mathrm{X}}(2)\frac{9}{12} = A, \\ f_{\mathrm{XY}}(3,7) =& f_{\mathrm{X}}(3)\frac{9}{12} = B. \end{aligned}$$

$A, B, f_{\mathrm{X}}(2), f_{\mathrm{X}}(3)$ を未知数として解くと, $A = \frac{6}{12}$, $B = \frac{3}{12}$.

L06-Q9

Quiz 解答:独立な確率変数の母期待値

1. X, Y は独立なので, $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$ であることに注意して, $\mathrm{E}[(-2X+3Y)(X+5Y)] = \mathrm{E}[-2X^2] + \mathrm{E}[-7XY] + \mathrm{E}[15Y^2] = -2(\mathrm{V}[X] + \mathrm{E}[X]^2) - 7\mathrm{E}[X]\mathrm{E}[Y] + 15(\mathrm{V}[Y] + \mathrm{E}[Y^2])$
2. 独立なので, $\mathrm{V}[XY] = \mathrm{V}[X] + \mathrm{V}[Y]$ であることに注意して, $\mathrm{V}[-2X+3Y] = \mathrm{V}[-2X] + \mathrm{V}[3Y] = 4\mathrm{V}[X] + 9\mathrm{V}[Y]$.

ここまで来たよ

復習+ちょっと I

L07-Q1

Quiz(離散的な確率変数の母平均値・母分散・母標準偏差・確率)

確率変数 X は次の確率分布に従う.

$$f_{10}(x) = \begin{cases} \frac{x}{55} & (0 \leq x \leq 10) \\ 0 & (\text{他}) \end{cases}$$

1. 確率 $P(X \leq 5)$ を求めよう.
2. 母平均値 $\mathrm{E}[X]$ を求めよう.
3. 母分散 $\mathrm{V}[X]$ を求めよう.

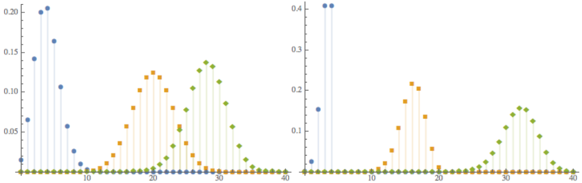
二項分布 高校 数学 B 西川確率統計定義 2.4(p.36)

二項分布

離散型確率変数 X が次の確率分布を持つとき, X はパラメタ n, p の二項分布 $\mathrm{B}(n,p)$ にしたがうという.

$$P(X=x)=f(x)=\begin{cases} {}_n\mathrm{C}_x\, p^x(1-p)^{n-x} & (x=0,1,2,3,\ldots,n) \\ 0 & (\text{他}) \end{cases}$$

意味: 確率 p で表の出るコインを n 回投げたとき, x 回表が出る確率.



$\mathrm{B}(40,0.1), \mathrm{B}(40,0.5), \mathrm{B}(40,0.7)$, $\mathrm{B}(4,0.8), \mathrm{B}(20,0.8), \mathrm{B}(40,0.8)$

二項分布の母平均値と母分散 西川確率統計例 2.2(p.44)

$$\mathrm{E}[X] = \boxed{}, \mathrm{V}[X] = \boxed{}$$

証明延期.

二項定理 高校 数学 A 西川確率統計定理 2.1(p.36)

$$(a+b)^n = \sum_{x=0}^{n} {}_n\mathrm{C}_x a^x b^{n-x}$$

$\mathrm{E}[1] =$

$$\boxed{}$$

L07-Q2

Quiz(二項分布)

確率 $p = \frac{2}{3}$ で表のでるコインを 100 回投げる.

1. 表が 40 回でる確率を求めよう. 階乗と巾乗と分数は簡単化・約分しなくてよい. それ以外の記号は使わないで答えること.
2. 表がでる回数の母平均値, 母分散を求めよう.

ベルヌーイ分布 西川確率統計定義 2.3(p.36)

ベルヌーイ分布

$n = 1$ の二項分布 $\mathrm{B}(1, p)$ のこと

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

意味:ベルヌーイ試行=(不公平な) コイン投げ．表がでる確率 p. 表 $x = 1$.

ベルヌーイ分布の母平均値と母分散 西川確率統計例 2.2(p.44)

$$\mathrm{E}[X] = \boxed{}, \qquad \mathrm{V}[X] = \boxed{}$$

ベルヌーイ分布と二項分布の関係 西川確率統計定理 2.2(p.37)

$X_1, X_2, \ldots, X_n$ が独立で $X_i \sim \mathrm{B}(1,p)$ のとき,
$U_n = X_1 + \cdots + X_n$ は $U_n \sim \mathrm{B}(n,p)$.

なぜなら

二項分布の母平均値と母分散の証明

(復習) 確率変数の和の母平均値と母分散

確率変数 $X_1, X_2, U_2 = X_1 + X_2$ を考える.
いつでも $\mathrm{E}[U_2] = \mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2]$.
X_1, X_2 が独立のとき $\mathrm{V}[U_2] = \mathrm{V}[X_1 + X_2] = \mathrm{V}[X_1] + \mathrm{V}[X_2]$.

L07-Q3

Quiz(ベルヌーイ分布)

ある宝くじは，あたりとはずれの 2 種類の結果だけがある．あたりの確率は 0.05 である．あたりの賞金は 1000 円，はずれの賞金は 0 円である．賞金を確率変数 Y (円) とする．

1. Y と，ベルヌーイ分布 $\mathrm{B}(1, p)$ に従う確率変数 X との関係を書こう．
2. Y の母平均値と母分散を求めよう．単位をつけよう．

ここまで来たよ

チェビシェフの不等式

チェビシェフの不等式 Chebyshev's inequality

西川確率統計定理 2.11(p.52)

X: 離散型 (または連続型) 確率変数
$\mu = \mathrm{E}[X]$: 母平均値
$\sigma^2 = \mathrm{V}[X]$: 母分散
$a > 0$: 任意の正の実数
のとき次が成立する.

$$P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$$

どんな X にも使えて便利な不等式. 意味は…

チェビシェフの不等式の証明

$P(|X-\mu| \geq a\sigma)$ を母期待値として書くと…

$$\begin{aligned}
P(|X-\mu| \geq a\sigma) &= \sum_x \mathbf{1}_{[|X-\mu|\geq a\sigma]}(x) \cdot f(x) \\
&\leq \sum_x \mathbf{1}_{[|X-\mu|\geq a\sigma]}(x) \frac{(x-\mu)^2}{(a\sigma)^2} f(x) \\
&\leq \sum_x \frac{(x-\mu)^2}{(a\sigma)^2} f(x) \\
&= \frac{1}{(a\sigma)^2} \sum_x (x-\mu)^2 \cdot f(x) \\
&= \frac{1}{(a\sigma)^2} \mathrm{V}[X] \\
&= \frac{1}{a^2}.
\end{aligned}$$

(連続型でも同様)

L07-Q4

西川確率統計演習 2.2(p.59)

独立同分布の性質 西川確率統計 §4.1

独立同分布 (i.i.d.) 西川確率統計 §4.1

離散型/連続型確率変数 $X_1, X_2, \ldots, X_n$ が, たがいに独立で, すべて同じ確率分布に従う (同じ確率関数 $f(x)$) とする.
これを $X_1, \ldots, X_n$ は独立同分布に従う (i.i.d.=independent and identically-distributed) という.

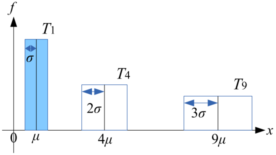
新しい確率変数: $U_n = X_1 + \cdots + X_n$
母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$ としたとき,

$$\mathrm{E}[U_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] = n \times \mu.$$

$$\mathrm{V}[U_n] = \sum_{i=1}^{n} \mathrm{V}[X_i] = n \times \sigma^2.$$

西川確率統計定理 4.2(p.85)

U_n の確率関数はこんな感じ?

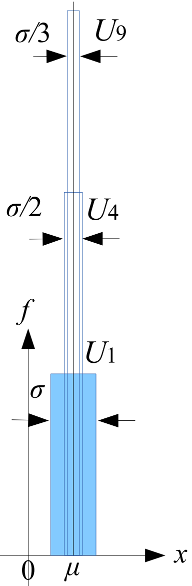


新しい確率変数: $W_n = \frac{1}{n}U_n = \frac{1}{n}(X_1 + \cdots + X_n)$

$$\mathrm{E}[W_n] = \mathrm{E}\left[\tfrac{1}{n}U_n\right] = \frac{1}{n} \times n \times \mu.$$

$$\mathrm{V}[W_n] = \mathrm{V}\left[\tfrac{1}{n}U_n\right] = \left(\frac{1}{n}\right)^2 \times n \times \sigma^2.$$

W_n の確率関数はこんな感じ?



L07-Q5

Quiz(独立同分布にしたがう変数の和)

確率変数 $X_1, \ldots, X_{100}$ は $\mathrm{E}[X_i] = 3, \mathrm{V}[X_i] = 7$ の独立同分布に従う.
次の確率変数の母平均値と母分散を求めよう.

1. 確率変数 $A = \frac{1}{100}(X_1 + X_2 + X_3 + \cdots + X_{100})$
2. 確率変数 $B = \frac{1}{10}(X_1 + X_2 + X_3 + \cdots + X_{100} - 100 \cdot 3)$
3. 確率変数 $C = \frac{1}{10\sqrt{7}}(X_1 + X_2 + X_3 + \cdots + X_{100} - 100 \cdot 3)$

大数の (弱) 法則 西川確率統計定理 4.1(p.84)

$X_1, \ldots, X_n$ が独立同分布にしたがい, $\mathrm{E}[X_i] = \mu$,
$\mathrm{V}[X_i] = \sigma^2$,$W = \frac{1}{n}\sum_{i=1}^{n} X_i$ のとき,

$$\forall \epsilon > 0 \lim_{n \to +\infty} P(|W_n - \mu| \geq \epsilon) = 0.$$

つまり n 大で W_n は $\mathrm{E}[W_n] = \mathrm{E}[X_i]$ に「必ず近い」(確率収束).
証明　$\mathrm{E}[W_n] = \mu, \mathrm{V}[W_n] = \sigma^2/n$ に注意して, W_n に対するチェビシェフの不等式を書くと,

$$P(|W_n - \mu| \geq a \times \frac{\sigma}{\sqrt{n}}) \leq \frac{1}{a^2}$$

$a = \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}$ とすると, $n \to +\infty$ で

$$P(|W_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n} \to 0$$

これが母平均値・母期待値の直観的意味. 要するに,

プチテスト計画

プチテスト 2017-11-22 水．1-609 実習室．30 ピーナッツ．
以下の出題計画は最終的なものではありません．2017-11-16 木に修正，確定します．一部の (数学的) 問題の PC(Moodle) による回答あり．Excel の使い方の問題は出題しません．

- データやグラフから平均値，分散，共分散，標準偏差，四分位数，四分位範囲などを求めその意味を解釈する (L02)
- データやグラフや平均値分散から標準得点，偏差値を求め，その意味を解釈する (文章題) (L03)
- データやグラフや平均値分散などから 相関係数，回帰係数，回帰直線を求めその意味を解釈する (L04，レポート)
- 1 次元の離散型確率変数について，確率分布から確率，母期待値，母平均値，母分散，母標準偏差を求める $\times n$ 問 (L05)
- 確率変数の 1 次式や 2 次式について，母平均値，母分散から，母平均値，母分散，母標準偏差を求める (L05,L06)
- 2 次元の離散型確率分布について，同時分布，周辺分布，母期待値，母分散，母共分散，独立性から母期待値，母共分散，母相関係数を求める (L06)
- 二項分布の確率，母平均値，母分散の式を利用して，確率変数の母平均値，母分散，母期待値を求める (L07)
- 独立同分布に従う確率変数 X_i の和 W について，X の情報から，W の母平均値，母分散を求める (L05,L07)
- 連続型確率変数について，確率密度関数から確率，母期待値，母平均値，母分散，母標準偏差を求める $\times n$ 問 (L08)

おすすめの準備方法: 出題範囲や方式は毎年変わるので，過去問題 (公開) 中心の準備はおすすめしません．上の出題計画を参照して，今年度の trial や準備問題を中心に準備することをおすすめします．Learn Math Moodle の準備問題は，点数はこれまでの最高点で固定されてますが，練習のための再受験が可能です．