二項分布, 独立同分布の和

樋口さぶろお https://hig3.net

龍谷大学理工学部数理情報学科

確率統計☆演習 L08(2020-11-16 Mon)

最終更新: Time-stamp: "2020-11-15 Sun 08:40 JST hig"

今日の目標

- 二項分布の母ナントカが求められる 岩薩林 確率・統計 §3.4
- 独立同分布の和の母ナントカが求められる
  岩薩林 確率・統計 例題 4.6(p.84)

L07-Q1

Quiz 解答:多変数の確率変数の期待値

1. $$p(x,y) = \begin{array}{|c||ccc|} \hline y\backslash x & 1 & 2 & 3 \\ \hline 0 & 0 & \frac{2}{12} & \frac{1}{12} \\ 2 & \frac{4}{12} & 0 & \frac{5}{12} \\ \hline \end{array}$$

$$g(x,y) = x^2 + \mathrm{e}^y = \begin{array}{|c||ccc|} \hline y\backslash x & 1 & 2 & 3 \\ \hline 0 & 1^2+\mathrm{e}^0 & 2^2+\mathrm{e}^0 & 3^2+\mathrm{e}^0 \\ 2 & 1^2+\mathrm{e}^2 & 2^2+\mathrm{e}^2 & 3^2+\mathrm{e}^2 \\ \hline \end{array}$$

$$\mathrm{E}[X^2+\mathrm{e}^Y] = (1^2+\mathrm{e}^0)0 + (2^2+\mathrm{e}^0)\tfrac{2}{12} + (3^2+\mathrm{e}^0)\tfrac{1}{12}$$
$$+(1^2+\mathrm{e}^2)\tfrac{4}{12} + (2^2+\mathrm{e}^2)0 + (3^2+\mathrm{e}^2)\tfrac{5}{12}$$

2. $$\mathrm{I}_{[XY\geq 2]}(X,Y) = \begin{array}{|c||ccc|} \hline y\backslash x & 1 & 2 & 3 \\ \hline 0 & 0 & 0 & 0 \\ 2 & 1 & 1 & 1 \\ \hline \end{array}$$

$$\mathrm{E}[\mathrm{I}_{[XY\geq 2]}(X,Y)] = 0\cdot 0 + 0\cdot\tfrac{2}{12} + 0\cdot\tfrac{1}{12} + 1\cdot\tfrac{4}{12} + 1\cdot 0 + 1\cdot\tfrac{5}{12} = \tfrac{9}{12}.$$

3.
$$p_{\mathrm{X}}(x) = \begin{cases} 4/12 & (x=1) \\ 2/12 & (x=2) \\ 6/12 & (x=3) \\ 0 & (\text{他}) \end{cases} \qquad p_{\mathrm{Y}}(y) = \begin{cases} 3/12 & (y=0) \\ 9/12 & (y=2) \\ 0 & (\text{他}) \end{cases}$$

4. (1 の別解) $\mathrm{E}[X^2 + \mathrm{e}^Y] = 2\mathrm{E}[X^2] + \mathrm{E}[\mathrm{e}^Y] =$ 周辺分布で計算.

L07-Q2

Quiz 解答:独立と限らない確率変数の母期待値

1. $\mathrm{E}[-2X + 3Y] \overset{\text{E5}}{=} \mathrm{E}[-2X] + \mathrm{E}[3Y] \overset{\text{E3}}{=} -2\mathrm{E}[X] + 3\mathrm{E}[Y] = 5.$
2. $\mathrm{V}[-2X + 3Y] \overset{\text{V1}}{=} \mathrm{E}[(-2X + 3Y)^2] - \mathrm{E}[-2X + 3Y]^2 \overset{\text{E4}}{=}$
$\mathrm{E}[(-2)^2X^2] + \mathrm{E}[2(-2)3XY] + \mathrm{E}[3^2Y^2] - (\mathrm{E}[-2X] + \mathrm{E}[3Y])^2 \overset{\text{E3,V1,C1}}{=}$
$(-2)^2\mathrm{V}[X] + 2(-2)(3)\mathrm{Cov}[X, Y] + 3^2\mathrm{V}[X] = 20 - 84 + 99 = 35.$

L07-Q3

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

1. 独立でない
2. $\mathrm{E}[X] = \frac{1}{7} \cdot 1 + \frac{6}{7} \cdot 3 = \frac{19}{7}$, $\mathrm{E}[Y] = \frac{3}{7} \cdot 2 + \frac{4}{7} \cdot 4 = \frac{22}{7}$,
3. $\mathrm{E}[XY] = \frac{1}{7} \cdot 2 \cdot 1 + \frac{2}{7} \cdot 3 \cdot 2 + \frac{4}{7} \cdot 3 \cdot 4 = \frac{62}{7}$.
   $\mathrm{Cov}[X, Y] = \frac{62}{7} - \frac{19}{7}\frac{22}{7} = \frac{16}{49}$. 独立性の必要条件が成立していないので, このことからも独立ではありえない.

L07-Q4

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

1. 独立である.
2. $\mathrm{E}[X] = 1 \cdot \frac{3}{12} + 3 \cdot \frac{9}{12} = \frac{30}{12}$. $\mathrm{E}[Y] = 2 \cdot \frac{4}{12} + 4\frac{8}{12} = \frac{46}{12}$
3. $\mathrm{E}[XY] \overset{\text{IE2}}{=} \frac{30}{12} \cdot \frac{46}{12}$. $\mathrm{Cov}[X, Y] \overset{\text{IC1}}{=} 0$.

4. $\mathrm{E}[X^2] = 1^2 \cdot \frac{3}{12} + 3^2 \cdot \frac{9}{12} = \frac{84}{12} = 7.\ \mathrm{E}[X^2 Y] \overset{\text{IE1}}{=} 7 \cdot \frac{46}{12}.$

L07-Q5
Quiz 解答:独立な確率変数の母期待値

1. X, Y は独立なので $\mathrm{E}[XY] \overset{\text{IE1}}{=} \mathrm{E}[X]\mathrm{E}[Y]$ であることに注意して,
$\mathrm{E}[(-2X+3Y)(X+5Y)] \overset{\text{E5}}{=} \mathrm{E}[-2X^2] + \mathrm{E}[-7XY] + \mathrm{E}[15Y^2] \overset{\text{V1,IE1}}{=}$
$-2(\mathrm{V}[X] + \mathrm{E}[X]^2) - 7\mathrm{E}[X]\mathrm{E}[Y] + 15(\mathrm{V}[Y] + \mathrm{E}[Y]^2) = 240.$
2. X, Y は独立なので $\mathrm{V}[g_1(X) + g_2(Y)] \overset{\text{IC2}}{=} \mathrm{V}[g_1(X)] + \mathrm{V}[g_2(Y)]$ であることに注意して,
$\mathrm{V}[-2X+3Y] \overset{\text{IC2}}{=} \mathrm{V}[-2X] + \mathrm{V}[3Y] \overset{\text{V2}}{=} (-2)^2\mathrm{V}[X] + 3^2\mathrm{V}[Y] = 119.$

ここまで来たよ

復習＋ちょっと（x で書かれた確率関数）I

L08-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)

整数に値をとる離散型確率変数 X は次の確率分布に従う．

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

1. 確率 $P(X \le 5)$ 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．

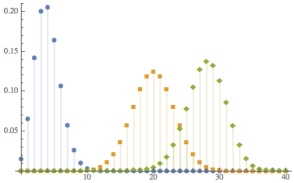
二項分布 高校 数学 B 岩薩林 確率・統計 §3.4

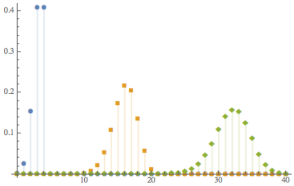
二項分布 岩薩林 確率・統計 (3.24)p.66

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．

$$p(x) = \begin{cases} \frac{n!}{x!(n-x)!}\, p^x (1-p)^{n-x} & (x = 0, 1, 2, 3, \ldots, n) \\ 0 & (\text{他}) \end{cases}$$

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．





$${}_n\mathrm{C}_x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

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

二項分布の確率関数のグラフ

$$p(x) = \begin{cases} \frac{n!}{x!(n-x)!} \, p^x (1-p)^{n-x} & (x = 0, 1, 2, 3, \ldots, n) \\ 0 & (\text{他}) \end{cases}$$

https://gallery.shinyapps.io/dist_calc/

https://shiny.rit.albany.edu/stat/binomial/

二項分布の母平均値と母分散 (証明延期) 岩薩林 確率・統計 定理 3.1(p.70)

$$\mathrm{E}[X] = \boxed{np}, \mathrm{V}[X] = \boxed{np(1-p)}$$

$\mathrm{E}[1] =$ 岩薩林 確率・統計 式 (3.26)–(3.27)

二項定理 高校 数学 A 岩薩林 確率・統計 (3.25)p.68

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

L08-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 $n!$ とべき乗 a^b と分数 $\frac{a}{b}$ は簡単化・約分しなくてよい．P,C などの記号は使わないで答えること．
2. 表がでる回数は確率変数である．母平均値，母分散を求めよう．

L08-Q3

Quiz(二項分布または独立同分布)

あるスーパーのおでんセット 1 パックには, 確率 $\frac{7}{10}$ で卵が 2 個, 確率 $\frac{3}{10}$ で卵が 0 個 (!) 入っている. おでんセットを 10 パックを買ったときに得られる卵の合計の個数を確率変数 Y とする. 各パックの卵の個数は独立とする.

1. 10 パック中 X パックが卵 2 個入りとする. X のしたがう確率分布と, X と Y の関係を答えよう.
2. 母平均値 $\mathrm{E}[Y]$ を求めよう.
3. 母分散 $\mathrm{V}[Y]$ を求めよう.
4. 確率 $P(Y = 16)$ を求めよう.

チェビシェフの不等式

チェビシェフの不等式 Chebyshev's inequality

岩薩林 確率・統計 §4.3(4.15)

X を離散型または連続型確率変数とする. $\mu = \mathrm{E}[X]$: 母平均値,
$\sigma^2 = \mathrm{V}[X]$: 母分散
$a > 0$: 任意の正の実数.
こ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. これが母平均値・母標準偏差のひとつの意味づけ.

$a > 1$ と思う. 母平均値 μ からの距離が母標準偏差 $\times a$ 以上離れた値が出る確率は, $1/a^2$ 以下. 母平均値から離れるほど, 確率は小さい. 母標準偏差 σ は離れ方の基準 (分布の幅).

チェビシェフの不等式の証明 (離散型)

意味のわからない $\mathrm{E}[\mathrm{I}_{[|X-\mu|\geq a\sigma]}(X)(X-\mu)^2]$ を定義に戻って書くと,

$$\begin{aligned}(a\sigma)^2 \times P(|X-\mu| \geq a\sigma) &= \sum_x \mathrm{I}_{[|X-\mu|\geq a\sigma]}(x)(a\sigma)^2\, p(x) \\ &\leq \sum_x \mathrm{I}_{[|X-\mu|\geq a\sigma]}(x)(x-\mu)^2\, p(x) \\ &\leq \sum_x (x-\mu)^2\, p(x) \\ &= \mathrm{V}[X] = \sigma^2\end{aligned}$$

連続型での証明 岩薩林 確率・統計 §4.3

ここまで来たよ

ベルヌーイ分布

ベルヌーイ分布

岩薩林 確率・統計 名前は出てこない

$n=1$ の二項分布 $\mathrm{B}(1,p)$ のこと

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

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

$$\mathrm{E}[X]=\boxed{p},\quad \mathrm{V}[X]=\boxed{p(1-p)}.\quad (\text{上の } f \text{ から直接計算できる})$$

確率変数の「同じ」

2 つの確率変数 $X = X_1, Y = X_2$.
$X = Y$

- 青色コイン 1 枚 ($\mathrm{B}(1, p)$) を 1 回投げた結果に, 2 つの変数名をつけた.
- X, Y は独立でない.

X, Y は同分布にしたがう

- 形の同じ青色コインと赤色コイン. 試行ごとの結果は異なる.
- $p_{\mathrm{X}}(x) = p_{\mathrm{Y}}(x)$. 母期待値も等しい. 例えば両方ともベルヌーイ分布にしたがい, $p_1 = p_2$.
- X, Y は独立でありうる (同時に投げて風に影響されたりしない限り).
  $\rightsquigarrow$ X, Y は独立同分布 (i.i.d.) にしたがう.

独立同分布 (i.i.d.)

岩薩林 確率・統計 定理 4.2(p.87) の仮定,p.113

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, すべて同分布に従う (同じ確率関数 $p(x)$ を持つ) とする.
このとき $X_1, \ldots, X_n$ は独立同分布に従う (i.i.d.=independent and identically-distributed) という.

ベルヌーイ分布と二項分布のもうひとつの関係

岩薩林 確率・統計 例題 4.6 の最初の 1 文

$X_1, X_2, \ldots, X_n$ が独立同分布 $\mathrm{B}(1,p)$ にしたがうとき,
$T_n = X_1 + \cdots + X_n$ は $T_n \sim \mathrm{B}(n,p)$.

なぜなら

X_i は 1 枚のコインのうち表が出た枚数. T_n は n 枚のコインのうち表が出た枚数.

(復習) 確率変数の和の母平均値と母分散

確率変数 $X_1, X_2, Y = aX_1 + bX_2$ を考える.
いつでも $\mathrm{E}[Y] = \mathrm{E}[aX_1 + bX_2] \overset{\text{E5,E3}}{=} a\mathrm{E}[X_1] + b\mathrm{E}[X_2]$.
X_1, X_2 が独立のとき $\mathrm{V}[Y] = \mathrm{V}[aX_1 + bX_2] \overset{\text{IC4}}{=} a^2\mathrm{V}[X_1] + b^2\mathrm{V}[X_2]$.

L08-Q4

Quiz(独立同分布の母期待値母分散)

確率変数 X_1, X_2, X_3 は独立同分布に従い, $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$ とする.
次の母期待値, 母分散を μ と σ^2 で表そう.

1. $\mathrm{E}[X_1 + X_1 + X_1]$
2. $\mathrm{E}[X_1 + X_2 + X_3]$
3. $\mathrm{E}[X_1 + X_1 + X_2]$
4. $\mathrm{V}[X_1 + X_1 + X_1]$
5. $\mathrm{V}[X_1 + X_2 + X_3]$
6. $\mathrm{V}[X_1 + X_1 + X_2]$
7. $\mathrm{E}[(X_1 + X_2 + X_3)^2]$

独立同分布にしたがう確率変数の和の性質 岩薩林 確率・統計 例題 4.6

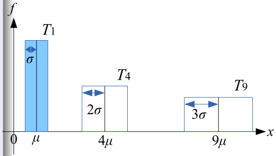
i.i.d にしたがう確率変数の和

$X_1, \ldots, X_n$: i.i.d. 母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$.
和の確率変数 $T_n = X_1 + \cdots + X_n$.

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

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

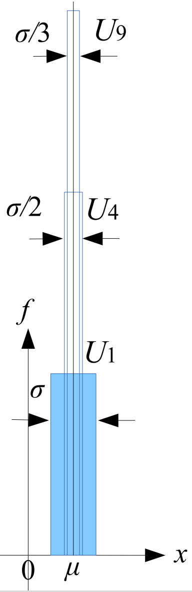
T_n の確率関数はこんな感じ?



二項分布の母平均値・母分散の延期していた証明

$X_1, \ldots, X_n \sim \mathrm{B}(1, p)$: i.i.d.
このとき, $T_n \sim \mathrm{B}(n, p)$ に対して,
$\mathrm{E}[T_n] = n\mathrm{E}[X_i] = n \times p, \mathrm{V}[T_n] = n \times \mathrm{V}[X_i] = np(1-p).$

U_n の確率関数はこんな感じ?



i.i.d にしたがう確率変数の和の $1/n$

$X_1, \ldots, X_n$: i.i.d.
新しい確率変数: $U_n = \frac{1}{n} T_n = \frac{1}{n}(X_1 + \cdots + X_n)$

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

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

L08-Q5

Quiz(独立同分布にしたがう確率変数の和)

ダーツで, 1 回投げるごとに点数が得られ, n 回投げた合計点で競うルールでプレイしている.
あるプレイヤーの i 回目の点数を確率変数 X_i とすると, X_i は独立同分布にしたがい, $\mathrm{E}[X_i] = 80, \mathrm{V}[X_i] = 200$ だという.
このプレイヤーが n 回投げた合計点を確率変数 $T_n = \sum_{i=1}^{n} X_i$ とする.

1. $\mathrm{E}[T_{10}]$, $\mathrm{V}[T_{10}]$ を求めよう.
2. $\mathrm{E}[\frac{1}{10}T_{10}]$, $\mathrm{V}[\frac{1}{10}T_{10}]$ を求めよう.
3. $\mathrm{E}[\frac{1}{100}T_{100}]$, $\mathrm{V}[\frac{1}{100}T_{100}]$ を求めよう.

大数の (弱) 法則アバウト版 岩薩林 確率・統計 §4.4

$X_1, \ldots, X_n$ が独立同分布にしたがい, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$, $U_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ のとき, n が十分大きいとき U_n は 'ほぼ μ に等しい' (U_n が μ から外れる確率はゼロに近づく)

(前と同じ問の別解)
L08-Q6

Quiz(二項分布または独立同分布)

あるスーパーのおでんセット 1 パックには, 確率 $\frac{7}{10}$ で卵が 2 個, 確率 $\frac{3}{10}$ で卵が 0 個 (!) 入っている. おでんセットを 10 パックを買ったときに得られる卵の合計の個数を確率変数 Y とする. 各パックの卵の個数は独立とする.

1. 10 パック中 X パックが卵 2 個入りとする. X のしたがう確率分布と, X と Y の関係を答えよう.
2. 母平均値 $\mathrm{E}[Y]$ を求めよう.
3. 母分散 $\mathrm{V}[Y]$ を求めよう.
4. 確率 $P(Y = 16)$ を求めよう.