二項分布

樋口さぶろお　https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

確率統計 I　L07(2021-05-26 Wed)

最終更新: Time-stamp: "2021-05-22 Sat 11:58 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
- 二項分布の文章題が解ける
- チェビシェフの不等式, 大数の法則, 母平均値・母分散の正確な意味が説明できる 岩薩林 確率・統計 §4.3,§4.4

L06-Q1

Quiz 解答:ベイズ推定

Y を色, X を当たり外れとすると,

$$
\begin{aligned}
&P(X=\text{当}\,|Y=\text{赤})\\
=&\frac{P(Y=\text{赤}\,|X=\text{当})P(X=\text{当})}{P(Y=\text{赤}\,|X=\text{当})P(X=\text{当})+P(Y=\text{赤}\,|X=\text{外})P(X=\text{外})}\\
=&\frac{\frac{1}{10}\frac{2}{10}}{\frac{1}{10}\frac{2}{10}+\frac{7}{10}\frac{8}{10}}=\frac{2}{58}.
\end{aligned}
$$

$Y\backslash X$	当たり	外れ
赤	$\frac{1}{10}\cdot\frac{2}{10}$	$\frac{7}{10}\cdot\frac{8}{10}$
白	$\frac{9}{10}\cdot\frac{2}{10}$	$\frac{3}{10}\cdot\frac{8}{10}$
合計	$\frac{2}{10}$	$\frac{8}{10}$

L06-Q2

L06-Q3

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，$\mathrm{V}[-2X+3Y] \overset{\text{IC2}}{=} \mathrm{V}[-2X] + \mathrm{V}[3Y] \overset{\text{V2}}{=} (-2)^2\mathrm{V}[X] + 3^2\mathrm{V}[Y] = 119.$

L06-Q4

Quiz 解答:独立同分布の母期待値母分散

1. 3μ.
2. 3μ.
3. 3μ
4. $9\sigma^2$.
5. $3\sigma^2$.
6. $5\sigma^2$.

7. $3(\sigma^2 + \mu^2) + 3 \times 2 \times \mu^2.$

L06-Q5

Quiz 解答:独立同分布にしたがう確率変数の和

1. $\mathrm{E}[S_{10}] = 800,\ \mathrm{V}[S_{10}] = 2000.$
2. $\mathrm{E}[\frac{1}{10}S_{10}] = 80,\ \mathrm{V}[\frac{1}{10}S_{10}] = 20.$
3. $\mathrm{E}[\frac{1}{100}S_{100}] = 80,\ \mathrm{V}[\frac{1}{100}S_{100}] = 2.$

ここまで来たよ

6 確率変数の独立性・独立同分布の和

7 二項分布
- 二項分布
- チェビシェフの不等式
- ベルヌーイ分布と二項分布・独立同分布の性質の応用
- 大数の弱法則

復習+ちょっと (x で書かれた確率関数) 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)

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.

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

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.

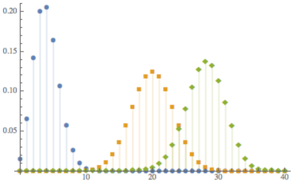
二項分布 高校 数学 B 岩薩林 確率・統計 §3.4

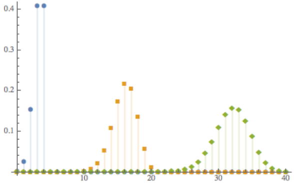
二項分布 岩薩林 確率・統計 (3.24)p.66

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

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

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 $n!$ 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.

L07-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.

ここまで来たよ

6 確率変数の独立性・独立同分布の和

7 二項分布
- 二項分布
- チェビシェフの不等式
- ベルヌーイ分布と二項分布・独立同分布の性質の応用
- 大数の弱法則

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

6. 確率変数の独立性・独立同分布の和

7. 二項分布
   - 二項分布
   - チェビシェフの不等式
   - ベルヌーイ分布と二項分布・独立同分布の性質の応用
   - 大数の弱法則

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 $x = 1$．

ベルヌーイ分布のモーメントと母平均値と母分散

$$\mathrm{E}[X^k] = \boxed{p}\ (k = 1, 2, \ldots), \quad \mathrm{V}[X] = \boxed{p(1-p)}.$$

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,
$S_n = X_1 + \cdots + X_n$ は $S_n \sim \mathrm{B}(n, p)$.

なぜなら

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

二項分布の母平均値・母分散の導出

$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{B}(1, p)$ とすると, $\mathrm{E}[X_i] = p, \mathrm{V}[X_i] = p(1-p)$.
このとき, 独立同分布の和 確率統計 I(2021)L06 $S_n = \sum_{i=1}^{n} X_i \sim \mathrm{B}(n, p)$ に対して,

$$\mathrm{E}[S_n] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n] \overset{\text{同分布}}{=} n \times p$$

$$\mathrm{V}[S_n] \overset{\text{独立}}{=} \mathrm{V}[X_1] + \cdots + \mathrm{V}[X_n] \overset{\text{同分布}}{=} n \times p(1-p).$$

(前と同じ問の別解)
L07-Q4

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.

L07-Q5

Quiz(二項分布または独立同分布)

あるくじは, 確率 0.1 で当たり, 確率 0.9 で外れる. 当たると賞金 7 円もらえ, 外れても賞金 2 円もらえる.
このくじを 100 回引くときの賞金の合計額を確率変数 X とする.

1. $X = 220$ となる確率を求めよう. 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. X の母平均値と母分散を求めよう (単位付きで).

ここまで来たよ

(復習) $\frac{1}{n}(X_1+\cdots+X_n)$ の分布

$X_1, X_2, \ldots, X_n$: i.i.d. のときの, $U_n = \frac{1}{n}(X_1+\cdots+X_n)$ の分布
L07-Q6

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 $S_n=\sum_{i=1}^{n} X_i$ とする.

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

大数の (弱) 法則

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, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$, $U_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ のとき, n が十分大きいとき U_n は 'ほぼ μ に等しい'

直観的意味

賭けは, たくさんの回数繰り返すと, 大もうけや大損の確率は減り, 決まった損得の額に集中する.

弱法則の正確な表現

U_n が μ から外れる確率は $n \to +\infty$ でゼロに近づく

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

証明 $X_1, \ldots, X_n$:i.i.d., $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$ なので,
$\mathrm{E}[U_n] = \mu, \mathrm{V}[U_n] = \mathrm{V}[X_i]/n = \sigma^2/n$.
U_n に対するチェビシェフの不等式より,

$$P\left(|U_n - \mu| \geq a \cdot \frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{a^2}.$$

$a > 0$ は任意なので, 不等式の右辺が ϵ になるように慎重に悪巧みして
$a = \frac{\epsilon\sqrt{n}}{\sigma}$ とおくと,

$$P\left(|U_n - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2 n}.$$

右辺は, $n \to +\infty$ で $\to 0$.

確率にまつわるいろいろな「収束」

大数の強法則

U_n の「近づく」(!) 先が μ でない確率はゼロである

$$P(|\lim_{n\to+\infty} U_n - \mu| > 0) = 0$$

確率にまつわる「収束」にはいろいろある…

- 概収束
- 確率収束
- 平均収束
- 法則収束
- $\vdots$