確率と統計　第 10 回

離散的な確率変数たち

6 月 22 日

Table of contents

日常的な変数のモデル化

日常の中で観測される多くの数値や記号は離散的な確率変数と思うことができる.

日常的な変数のモデル化

日常の中で観測される多くの数値や記号は離散的な確率変数と思うことができる.

例えば, ある区で一日に発生する交通事故の件数 X を記録するとしよう.

日常的な変数のモデル化

日常の中で観測される多くの数値や記号は離散的な確率変数と思うことができる.

例えば, ある区で一日に発生する交通事故の件数 X を記録するとしよう. このとき X をどのような分布を持つ確率変数としてモデル化すればよいだろうか.

ベルヌーイ分布: 成功か失敗か

ある保険会社の保険制度には免責条項があり，ある金額以下では被保険者個人の全額負担となり，その金額を超える場合には個人と会社で分担して負担することになる．

ベルヌーイ分布: 成功か失敗か

ある保険会社の保険制度には免責条項があり，ある金額以下では被保険者個人の全額負担となり，その金額を超える場合には個人と会社で分担して負担することになる．

そこで個人の全額負担となる場合を会社にとっての "成功" とし，分担負担となる場合を "失敗" とする．

ベルヌーイ分布: 成功か失敗か

ある保険会社の保険制度には免責条項があり，ある金額以下では被保険者個人の全額負担となり，その金額を超える場合には個人と会社で分担して負担することになる．

そこで個人の全額負担となる場合を会社にとっての“成功”とし，分担負担となる場合を“失敗”とする．

このとき，成功 or 失敗は二値のみをとる確率変数とみなせる．

ベルヌーイ分布: 成功か失敗か

成功確率を $p = 0.7$ としよう.

ベルヌーイ分布: 成功か失敗か

成功確率を $p = 0.7$ としよう.

成功を 1 とし, 失敗を 0 とすると, 例えば 10 回試行を観測したときの標本比率

$$
\begin{aligned}
\hat{p} &= \frac{1+1+0+1+0+0+0+0+1+1}{10} \\
&= 0.5
\end{aligned}
$$

は成功確率 p の近似値になる.

ベルヌーイ分布: 成功か失敗か

定義

二つの値のみをとる確率変数

$$X = \{0, 1\}$$

を**ベルヌーイ確率変数**という.

ベルヌーイ分布: 成功か失敗か

定義

二つの値のみをとる確率変数

$$X = \{0, 1\}$$

をベルヌーイ確率変数という. 1 をとる確率が p で, 0 をとる確率が $1 - p$ ならば平均と標準偏差は

$$\mu = p, \quad \sigma = \sqrt{p(1-p)}$$

となる.

ベルヌーイ試行になる条件

以下の試行はベルヌーイ試行になるだろうか.

(a) 箱の中からくじを順番に引いていき, 当たりか外れかを観測する.

(b) サイコロを転がして出た目.

ベルヌーイ試行になる条件

以下の試行はベルヌーイ試行になるだろうか.

(a) 箱の中からくじを順番に引いていき, 当たりか外れかを観測する.

(b) サイコロを転がして出た目.

実はどちらもベルヌーイ試行にはならない. (a) では各試行が独立ではなく, (b) では出る目の値が 6 通りある.

幾何分布: 初めて成功するまで

目玉焼きを作ることを考える．一回の試行で成功する確率は 0.7 とし，各試行は独立とする．このときの初めて成功するまでの回数 n を考える

幾何分布: 初めて成功するまで

目玉焼きを作ることを考える．一回の試行で成功する確率は 0.7 とし，各試行は独立とする．このときの初めて成功するまでの回数 n を考える

回数 n は値 $n = 1, 2, \cdots$ をとり，その確率分布は

$$P(n) = (1 - 0.7)^{n-1} \times 0.7$$

となる．このように確率は回数 n の増大に対して指数的な速さで減少する．

幾何分布: 初めて成功するまで

定義

成功の確率が p である独立なベルヌーイ試行を繰り返すとき, 初めて成功するまでの回数 n が従う確率分布を**幾何分布**という:

幾何分布: 初めて成功するまで

定義

成功の確率が p である独立なベルヌーイ試行を繰り返すとき, 初めて成功するまでの回数 n が従う確率分布を幾何分布という:

$$P(n) = (1-p)^{n-1}p, \quad n = 1, 2, \cdots,$$

$$\mu = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}, \quad \sigma = \sqrt{\frac{1-p}{p^2}}.$$

幾何分布になるための条件

初めて成功するまでの回数 n が幾何分布に従うためには, 各試行が独立で同一のベルヌーイ分布でなければならない.

幾何分布になるための条件

初めて成功するまでの回数 n が幾何分布に従うためには, 各試行が独立で同一のベルヌーイ分布でなければならない.

例えば, 目玉焼きの成功か失敗かをみる試行で徐々に上達していくとすれば各試行は独立ではないから, 初めて成功するまでの回数は幾何分布に従わない.

二項分布: 何回成功するか

同じく目玉焼きを作る試行を考え, 成功率は 0.7 で各試行は独立とする. n 回作るとき, そのうち k 回成功する確率を考えると

二項分布: 何回成功するか

同じく目玉焼きを作る試行を考え, 成功率は 0.7 で各試行は独立とする. n 回作るとき, そのうち k 回成功する確率を考えると

$$P(k) = {}_nC_k \times 0.7^k \times (1-0.7)^{n-k}$$

となる $(k = 0, 1, \cdots, n)$.

二項分布: 何回成功するか

定義

成功率が p の独立で同一なベルヌーイ試行を n 回繰り返す．このときの成功する回数 k が従う分布を**二項分布**という:

二項分布: 何回成功するか

定義

成功率が p の独立で同一なベルヌーイ試行を n 回繰り返す. このときの成功する回数 k が従う分布を二項分布という:

$$P(k) = {}_nC_k p^k (1-p)^{n-k}$$

$$\mu = np, \quad \sigma^2 = np(1-p), \quad \sigma = \sqrt{np(1-p)}.$$

二項分布: 何回成功するか

二項分布の平均や分散, 標準偏差は以下のようにしても導出できる.

二項分布: 何回成功するか

二項分布の平均や分散, 標準偏差は以下のようにしても導出できる.

ベルヌーイ確率変数 $X = \{0, 1\}$ の n 個の独立なコピー

$$X_1, X_2, \ldots, X_n$$

を考える.

二項分布: 何回成功するか

二項分布の平均や分散, 標準偏差は以下のようにしても導出できる.

ベルヌーイ確率変数 $X = \{0, 1\}$ の n 個の独立なコピー

$$X_1, X_2, \ldots, X_n$$

を考える. このとき, 成功回数 k は

$$S = X_1 + X_2 + \cdots + X_n$$

で与えられる.

二項分布: 何回成功するか

すると独立な確率変数 $X_1, X_2, \ldots, X_n$ の線形結合の性質から

二項分布: 何回成功するか

すると独立な確率変数 $X_1, X_2, \ldots, X_n$ の線形結合の性質から

$$E(S) = nE(X) = np,$$
$$V(S) = nV(X) = np(1-p).$$

これで二項分布の平均と分散がベルヌーイ変数 X から求まった.

二項モデルが適用できる状況

以下の状況で二項モデルを使うことができる.

二項モデルが適用できる状況

以下の状況で二項モデルを使うことができる.

(a) 試行は独立である.
(b) 試行回数 n は固定されている.
(c) それぞれの試行の結果は成功か失敗のどちらかである.
(d) 成功確率 p はすべての試行で一定.

二項モデルが適用できる状況

以下の状況で二項モデルを使うことができる.

(a) 試行は独立である.
(b) 試行回数 n は固定されている.
(c) それぞれの試行の結果は成功か失敗のどちらかである.
(d) 成功確率 p はすべての試行で一定.

例題: ある区での一日の事故発生件数 k は二項モデルに従わない. (a)-(d) のどれが満たされないか.

正規分布での近似

二項分布は標本サイズ n が十分に大きく, 以下の成功・失敗条件を満たすとき, 同じ平均と標準偏差を持つ正規分布 $N(\mu, \sigma)$ で近似することができる:

正規分布での近似

二項分布は標本サイズ n が十分に大きく, 以下の成功・失敗条件を満たすとき, 同じ平均と標準偏差を持つ正規分布 $N(\mu, \sigma)$ で近似することができる:

$$np \geq 10 \quad \text{かつ} \quad n(1-p) \geq 10.$$

負の二項分布: 何回目で成功するか 2

再び目玉焼きを作る試行を考える. 各試行は独立とし, 成功確率は 0.7 とする. このときはじめて k 回成功するまでの試行回数 n の分布を考える. 確率分布は以下のようになる:

負の二項分布: 何回目で成功するか 2

再び目玉焼きを作る試行を考える. 各試行は独立とし, 成功確率は 0.7 とする. このときはじめて k 回成功するまでの試行回数 n の分布を考える. 確率分布は以下のようになる:

$$\begin{aligned} P(n) &= {}_{n-1}C_{k-1} \times 0.7^{k-1} \times (1-0.7)^{(n-1)-(k-1)} \times 0.7 \\ &= {}_{n-1}C_{k-1} \times 0.7^{k} \times (1-0.7)^{n-k}. \end{aligned}$$

ここで $n = k, k+1, \ldots$

負の二項分布: 何回目で成功するか 2

定義

成功率 p の独立なベルヌーイ試行を行うとき, 初めて k 回成功する回数 n が従う分布を負の二項分布という:

$$P(n) = {}_{n-1}C_{k-1} p^k (1-p)^{n-k}.$$

ポアソン分布: 大集団での稀な出来事

ある区での一日の交通事故発生件数を k とする.

ポアソン分布: 大集団での稀な出来事

ある区での一日の交通事故発生件数を k とする.

このような各人がほぼ独立で大きな母集団で起る稀なイベントの数は以下の確率分布に従う:

ポアソン分布: 大集団での稀な出来事

ある区での一日の交通事故発生件数を k とする.

このような各人がほぼ独立で大きな母集団で起る稀なイベントの数は以下の確率分布に従う:

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

ここで $\lambda > 0$ は平均発生件数を表す. ゆえに $\mu = \lambda$. さらに $\sigma = \sqrt{\lambda}$ となる. このような分布をポアソン分布とよぶ.