

表の出る確率 $p = \frac{5}{6}$, 試行回数 $n = 6480$ の二項分布に従う確率変数が $[5372, 5413]$ に入る確率 $P(5372 \leq X \leq 5413)$ を正規分布で近似して求めよ。ただし、標準正規分布の確率を計算する場合、変換後の両端の値は小数第3位を四捨五入した値を用いること。

二項分布の母平均 μ と母分散 σ^2 を求めると

$$\mu = np = 6480 \times \frac{5}{6} = 5400$$

$$\sigma^2 = np(1-p) = 6480 \times \frac{5}{6} \times \left(1 - \frac{5}{6}\right) = 900 = 30^2$$

であるから、正規分布 $N(5400, 30^2)$ で近似できる。従って

$$Z = \frac{X - 5400}{30}$$

で変換すれば標準正規分布 $N(0, 1^2)$ になる。

よって

$$\begin{aligned} P(5372 \leq X \leq 5413) &= P\left(\frac{5372 - 5400}{30} \leq Z \leq \frac{5413 - 5400}{30}\right) \\ &= P\left(-\frac{14}{15} \leq Z \leq \frac{13}{30}\right) = P(-0.9333\cdots \leq Z \leq 0.4333\cdots) \\ &= P(-0.93 \leq Z \leq 0.43) \\ &= P(-0.93 \leq Z \leq 0) + P(0 < Z \leq 0.43) \\ &= 0.32381 + 0.16640 = 0.49021 = 0.490 \end{aligned}$$

二項分布の確率を精密に計算すると $0.50161\cdots$ であるから、ある程度近い値であることがわかる。さらに半整数補正を行い近似の精度を良くするためには、区間の幅を上限下限とも 0.5 ずつ範囲を広げればよい。つまり、確率を求める範囲を $[5351, 5428]$ から $[5371.5, 5413.5]$ にすることで

$$\begin{aligned} P(5371.5 \leq X \leq 5413.5) &= P\left(\frac{5371.5 - 5400}{30} \leq Z \leq \frac{5413.5 - 5400}{30}\right) \\ &= P(-0.95 \leq Z \leq 0.45) \\ &= P(-0.95 \leq Z \leq 0) + P(0 < Z \leq 0.45) \\ &= 0.32894 + 0.17364 = 0.50258 = 0.503 \end{aligned}$$

のように、真の値である $0.50161\cdots$ にかなり近い値が得られる。

・最尤推定量

指数分布 $f(x|\lambda) = \lambda e^{-\lambda x}$ の尤度関数および対数尤度関数を求め、パラメータ λ の最尤推定量を求めよ。

n 個の標本値 x_1, x_2, \dots, x_n が得られたときの尤度関数は

$$L(\lambda|x_1, x_2, \dots, x_n) = (\lambda e^{-\lambda x_1}) \times (\lambda e^{-\lambda x_2}) \times \dots \times (\lambda e^{-\lambda x_n}) = \lambda^n \times e^{-\lambda \sum x_i}$$

であり、対数尤度関数は

$$\log L(\lambda|x_1, x_2, \dots, x_n) = n \log \lambda - \lambda \sum_{i=1}^n x_i$$

となる。対数尤度関数を λ で微分して最大となる点を求めると

$$\frac{d}{d\lambda} \log L(\lambda|x_1, x_2, \dots, x_n) = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0$$

$$\Rightarrow \frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow \frac{1}{\lambda} = \bar{x} \Rightarrow \lambda = \frac{1}{\bar{x}}$$

つまり λ の最尤推定量は $\hat{\lambda} = \frac{1}{\bar{x}}$ である。

ちなみに、指数分布の母平均は $\frac{1}{\lambda}$ なので、指数分布の場合も標本平均 \bar{x} は母平均 μ の不偏推定量かつ最尤推定量になっている。

《尤度関数のまま微分する場合》

$$\begin{aligned} \frac{d}{d\lambda} L(\lambda|x_1, x_2, \dots, x_n) &= n\lambda^{n-1} \times e^{-\lambda \sum x_i} - \left(\sum_{i=1}^n x_i \right) \times \lambda^n \times e^{-\lambda \sum x_i} \\ &= \left(n - \lambda \sum_{i=1}^n x_i \right) \times \lambda^{n-1} \times e^{-\lambda \sum x_i} \end{aligned}$$

この値を 0 にするには $\lambda^{n-1} > 0, e^{-\lambda \sum x_i} > 0$ なので $n - \lambda \sum_{i=1}^n x_i = 0$ である。

$$\text{よって } \lambda \sum_{i=1}^n x_i = n \Rightarrow \lambda = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i} = \frac{1}{\bar{x}} \quad \text{と同じ結果になる。}$$

資料置場

<https://www.gen.kanagawa-it.ac.jp/takeda/class/>

確率統計 母平均 μ の区間推定

・平均の区間推定（母分散既知）

母集団が $N(\mu, \sigma^2)$ に従うとき、無作為抽出された n 個のデータを使って計算された標本平均は $N(\mu, \frac{\sigma^2}{n})$ に従うことが理論的にわかっている。よって

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

は標準正規分布 $N(0, 1)$ に従う。そこで

$$P(-z_\alpha < Z < z_\alpha) = 1 - \alpha \quad (1 - \alpha \text{ は事前に決めた確率})$$

になるように z_α を決める。確率の括弧の中身は

$$-z_\alpha < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_\alpha \quad \Rightarrow \quad \bar{X} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

と変形できるので、確率変数 \bar{X} を実際の値 \bar{x} で置き換えて $[\bar{x} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}]$ の間に真の平均 μ が入っている確率は、事前に決めた確率となる。

$P(-z_\alpha < Z < z_\alpha) = 1 - \alpha$ となる z_α の求め方については、分布の両端に $\frac{\alpha}{2}$ となるように決めればよいので、

$$P(0 < Z < z_\alpha) = \frac{1}{2} - \frac{\alpha}{2}$$

となる z_α を探せばよい。例えば $\alpha = 0.05$ (95%) の場合

$$P(0 < Z < z_\alpha) = 0.5 - 0.025 = 0.475$$

なので、標準正規分布の表の確率が 0.475 に一番近い値を探せばよい。 $z = 1.96$ のとき 0.4750 であるから、母平均の 95%信頼区間は

$$\bar{x} - 1.96 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

となる。良く使う $\alpha = 0.05$ と $\alpha = 0.01$ のときの z_α の正確な値は次の通りである。

$$\alpha = 0.05 \text{ のとき } z_\alpha = 1.95996 \dots = 1.960$$

$$\alpha = 0.01 \text{ のとき } z_\alpha = 2.57582 \dots = 2.576$$

これらの値は前に配布した標準正規分布の表の下に確率の値から z_α の値を求める表を載せてあります。試験の際には同じものを配布しますので、使い方を覚えてください。

中間試験は 11月22日（金）です。

確率統計 S 関数電卓の使い方

次の9つのデータの標本平均 \bar{x} , 標本分散 s^2 , 標本標準偏差 s を求めよ。

33, 54, 36, 42, 39, 47, 51, 40, 45

1) 関数電卓を統計モードに変更

SHARP: MODE キー → STAT → SD の順

CASIO: MODE キー → STAT → 1-VAR の順

(人数が多かった機種の一例です)

統計モードの場合、電卓のどこかに「STAT」や「SD」も文字が入ります。

ちなみに元の電卓モードに戻す場合は

SHARP: MODE キー → COMP もしくは NORMAL の順

CASIO: MODE キー → COMP の順

2) データの入力

基本的には データ入力 ⇒ 確定ボタン の順を繰り返す

確定ボタンはほとんどの機種が「M+」を押すだけ!!

(M+の下に《DATA》,《DT》と書いてあります)

3) 結果の出力

標本平均 \bar{x}

SHARP: ALPHA キー → \bar{x} と書いてある数字 → = の順

CASIO: SHIFT キー → STAT → VAR → \bar{x} → = の順

標本標準偏差 s

SHARP: ALPHA キー → sx と書いてある数字 → = の順

CASIO: SHIFT キー → STAT → VAR → $x\sigma_{n-1}$ → = の順

分散 s^2 の値は標本標準偏差 s を2乗すればよい。

以上のことを踏まえて標本平均 \bar{x} 、標本標準偏差 s 、標本分散 s^2 の値を計算すると

$$\bar{x} = 43 \quad (= 43.0)$$

$$s = 6.89202\dots \quad (= 6.89)$$

$$s^2 = 47.5 \quad (= 47.5)$$

となります。

・関数電卓の使い方

次の9個のデータの標本平均・標本分散を求めよ。

33, 54, 36, 42, 39, 47, 51, 40, 45

解答)

$$\bar{x} = \frac{33 + 54 + 36 + 42 + 39 + 47 + 51 + 40 + 45}{9} = \frac{387}{9} = 43$$

$$s^2 = \frac{1}{9-1} \sum_{i=1}^9 (x_i - 43)^2 = \frac{100 + 121 + 49 + 1 + 16 + 16 + 64 + 9 + 4}{8} = \frac{95}{2} = 47.5$$

$$s = \sqrt{47.5} = 6.89202 \dots = 6.89$$

参考) 標本分散の通常電卓を使った計算方法

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

であるから、データの2乗和 $\sum_{i=1}^n x_i^2$ を計算できれば良い。

2乗和の計算は電卓のメモリ機能を使えば簡単にできる。

$$\sum_{i=1}^n x_i^2 = 33^2 + 54^2 + 36^2 + 42^2 + 39^2 + 47^2 + 51^2 + 40^2 + 45^2 = 17021$$

$$s^2 = \frac{1}{9-1} \left(17021 - 9 \times 43^2 \right) = \frac{17021 - 16641}{8} = \frac{380}{8} = \frac{95}{2} = 47.5$$

のように定義通り計算した結果と一致する。

・母分散が既知の場合の区間推定

母集団が母分散 $\sigma^2 = 2.75^2$ の正規分布に従うとき、9つのデータ

34.7, 39.2, 42.9, 40.1, 36.6, 41.5, 35.2, 39.2, 38.0

が与えられたとき。

- (1) 関数電卓を使い標本平均 \bar{x} 及び標本分散 s^2 を求めよ。
- (2) 母分散が既知の場合の99%信頼区間を求めよ。

2019年度神奈川工科大学 確率統計S 演習問題	学科	学年	組	学 籍 番 号	氏 名	

提出先：K3-3309号室前 18番のボックス 提出期限：11月11日（月）17時頃まで