確率統計　前回の解答　　第３回分

１）区間 $[1,6]$ の一様分布において、区間 $[1.25, 3.75]$ に入る確率。

一様分布の確率密度関数は $f(x) = \dfrac{1}{6-1} = \dfrac{1}{5}$ $(1 \leqq x \leqq 6)$ であるから

$$P(1.25 \leqq X \leqq 3.75) = \int_{1.25}^{3.75} \frac{1}{5}\,dx = \left[\frac{x}{5}\right]_{1.25}^{3.75} = \frac{3.75-1.25}{5}$$

$$= \frac{2.50}{5} = 0.500 \left(= \frac{1}{2}\right)$$

注）有効桁数に注意

２）パラメータ $\lambda = \frac{1}{2}$ の指数分布において、区間 $[2.00, 8.00]$ に入る確率。

指数分布の確率密度関数は $f(x) = \frac{1}{2}e^{-\frac{1}{2}x}$ $(0 \leqq x)$ であるから

$$P(2.00 \leqq X \leqq 8.00) = \int_{2.00}^{8.00} \frac{1}{2}e^{-\frac{1}{2}x}\,dx = \left[-e^{-\frac{1}{2}x}\right]_{2.00}^{8.00} = -e^{-4.00} - \left(-e^{-1.00}\right)$$

$$= e^{-1.00} - e^{-4.00} = 0.349563\cdots = 0.350$$

注）解答は $e^{-1} - e^{-4}$ でも 0.350 でも可

３）標準正規分布 $N(0,1)$ において、区間 $[-2.28, -0.40]$ に入る確率。

既に標準正規分布なので特に変換をする必要なく表を使うことができる。

$$P(-2.28 \leqq X \leqq -0.40) = P(0.40 \leqq X \leqq 2.28) \quad \text{（分布が左右対称）}$$

$$= P(0 \leqq X \leqq 2.28) - P(0 \leqq X < 0.40) \quad \text{（0 からの確率にして引き算）}$$

$$= 0.48870 - 0.15542 = 0.33328 = 0.333 \quad \text{（表を使って確率の計算）}$$

注１）確率は負にならない（０から１の間）。

区間が負の範囲であっても確率は正の値になる（分布が左右対称であるため）。

$$P(-2.37 \leqq X \leqq 0) = P(0 \leqq X \leqq 2.37) = 0.49111$$

注２）標準正規分布の確率計算

区間に 0 を含む場合　　$\Rightarrow$ 0 で 2 つに分割して、それぞれの和を計算する

区間で考えると $[-1,2] = [-1,0] + [0,2]$ なので

$$P(-1 \leqq Z \leqq 2) = P(-1 \leqq Z \leqq 0) + P(0 \leqq Z \leqq 2)$$

資料（主に解答と演習問題）の置き場

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

演習の紙をなくした場合は、上記ページからダウンロードして印刷し、提出すること。

確率統計 期待値と平均・分散　　第４回

確率変数 X の期待される値として期待値は次のように定義されている。

$$E[X] = \begin{cases} \displaystyle\sum_i x_i \times P(X = x_i) & \text{（離散型）} \\ \displaystyle\int_{-\infty}^{\infty} x f(x)\, dx & \text{（連続型）} \end{cases}$$

この値は、平均値とも呼ばれ、一般に μ を使って表すことが多い。

さらに一般に確率変数 X を使った関数 $h(X)$ の期待値は次のように定義されている。

$$E[h(X)] = \begin{cases} \displaystyle\sum_i h(x_i) \times P(X = x_i) & \text{（離散型）} \\ \displaystyle\int_{-\infty}^{\infty} h(x) \times f(x)\, dx & \text{（連続型）} \end{cases}$$

また、分布の平均からのばらつき具合を表す統計量として、分散 σ^2 が定義されているがこの値は期待値を使って、$E[(X-\mu)^2]$ と定義している。

例１．サイコロの平均・分散

サイコロは１，２，３，４，５，６の目がそれぞれ確率 $\dfrac{1}{6}$ で出現する離散型である。よって、平均値 μ は

$$\mu = E[X] = \sum_{x=1}^{6} xP(X = i) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5$$

また、分散の値は

$$\sigma^2 = E[(X-\mu)^2] = \sum_{x=1}^{6} \left(x - \frac{7}{2}\right)^2 P(X = i)$$

$$= \left(1 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \cdots + \left(6 - \frac{7}{2}\right)^2 \times \frac{1}{6} = \frac{35}{12} = 2.9166\cdots$$

例２．区間 $[a, b]$ の一様分布の平均・分散

確率密度関数は $f(x) = \dfrac{1}{b-a}$ であることから、平均値 μ は

$$\mu = E[X] = \int_a^b x \times \frac{1}{b-a}\, dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

さらに分散 σ^2 は

$$\sigma^2 = E[(X-\mu)^2] = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \times \frac{1}{b-a}\, dx = \left[\frac{(a+b-2x)^3}{24(b-a)}\right]_a^b$$

$$= \frac{(a-b)^3 - (b-a)^3}{24(a-b)} = \frac{(b-a)^2}{12}$$

なお、分散の値は次のように計算することもできる。

$$E[(X-\mu)^2] = E[X^2] - \mu^2$$

平均と分散の推定（母集団と標本）

この授業の前半では確率論の講義として、様々な分布や平均・分散及び期待値等の計算を行ってきた。しかしながら、これらの結果はすべて理論的な話であって、現実の世界でこれから起こる事象の確率が事前にわかっていることはほとんど無い（たとえサイコロであっても本当にすべての出る目が等しく $\frac{1}{6}$ であることの保障は無い）。後半の統計学では、実験などによって得られるデータから未知である分布を推定することや、仮説が正しいかどうかを判断すること等を行う。もちろん得られるデータはなんらかの確率分布に従っているため、実験を行うたびに異なるデータが得られる。したがって、データから推定される値も実験ごとに異なる数値になる。統計学は確率論を使って、できるだけ正しい結果になるように推定する方法を考える。

母平均と標本平均・母分散と標本分散

母集団から確率変数 X が一つ取り出されたとき、X の取り得る値として期待される値、いわゆる平均値 μ は $\mu = E[X]$ で定義され、平均値からのバラツキ具合を表す分散 σ^2 は平均からの差の 2 乗の期待値 $\sigma^2 = E[(X-\mu)^2]$ で定義されている。

平均、分散共に母集団の確率（もしくは確率密度関数）が必要になるので、実際には未知であることが多い。そこで母集団から得られた n 個の標本 $x_1, x_2, \cdots, x_n$ を使って、次のように平均 μ，分散 σ^2 を $\bar{x}$ と s^2 を使って推定する。

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

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

$$\left(= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \right)$$ 【教科書によってはこちらを使っている】

一般に、母集団から計算された平均・分散を母平均 (μ)・母分散 (σ^2) と呼び、標本から計算された平均・分散を標本平均 $(\bar{x})$・標本分散 (s^2) と呼び、記号も区別をしている。また、分散の値の平方根（つまり σ もしくは s）のことを標準偏差と呼ぶ。

例題. 下記の表のような確率をもつ離散型の確率分布の母平均を求めよ。

X	-3	0	7	9
確率	$\frac{4}{25}$	$\frac{1}{5}$	$\frac{6}{25}$	$\frac{2}{5}$

この問題は母平均 μ を求めるので、定義どおり計算する必要がある。

$$\mu = E[X] = (-3)\times\frac{4}{25} + (0)\times\frac{1}{5} + (7)\times\frac{6}{25} + (9)\times\frac{2}{5} = \frac{24}{5} = 4.80 \qquad \left(\sigma^2 = \frac{564}{25} = 22.56\right)$$

これを次のように標本平均の計算をする人がいるが、もちろん間違いである。

$$\frac{1}{4}\times(-3+0+7+9) = \frac{13}{4} = 3.25$$

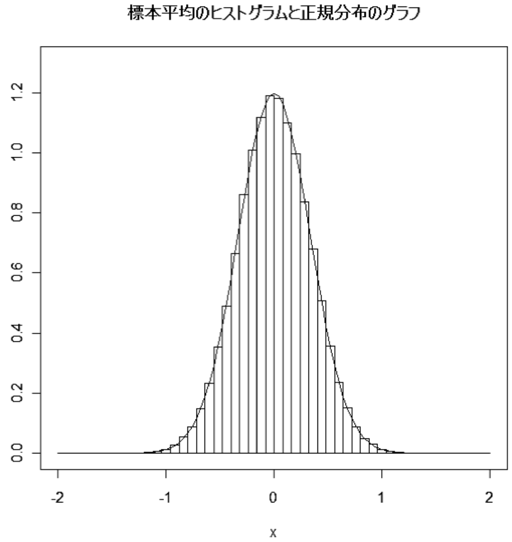
注：$-3, 0, 7, 9$ が 1 回ずつ出現したわけではないので、上記のような計算はもともとできない

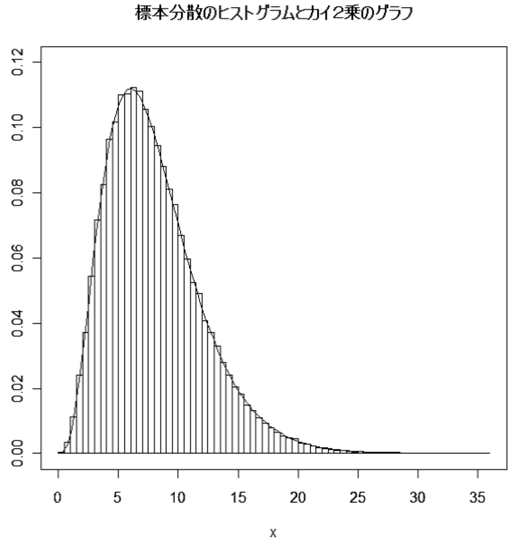
標本平均と標本分散の分布

母集団の分布が平均 μ，分散 σ^2 の正規分布 $N(\mu,\sigma^2)$ に従うとき、n 個の標本 $x_1, x_2, \cdots, x_n$ が得られたとき、標本平均と標本分散は次のような分布に従う。

・標本平均 $\bar{x}$ は平均は同じ μ，分散は n で割った $\frac{\sigma^2}{n}$ の正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従う
・標本分散 s^2 は $\frac{n-1}{\sigma^2}s^2$ で変換すると自由度 $(n-1)$ の χ^2 分布に従う。

例．標準正規分布から標本 $x_1, \cdots, x_9$ を得た場合のヒストグラムと理論分布





標本平均の分散は n が大きくなると小さくなるので、推定がよくなることがわかる（分散が小さいと μ の近くにしか出現しない）。そのため、実験などではなるべく多くのデータを得るように努力をしている。標本平均だけでなく、一般にパラメータの推定はデータの個数が多いほうが良い。

2019年度 確率統計　小テスト解答用紙　2019.10.15

・母平均と母分散

下記の表のような確率をもつ離散型の確率分布の母平均と母分散を求めよ。

X	-4	1	3	11
確率	$\frac{1}{10}$	$\frac{9}{20}$	$\frac{1}{5}$	$\frac{1}{4}$

なお、母分散は $E[(X-\mu)^2] = E[X^2] - \mu^2$ を用いて計算しても良い。

（裏面にもう１問あります）

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

提出先：K3-3309号室前　17番のボックス　提出期限：１０月１６日（水）１７時頃まで

・標本平均

・実際にコイン投げを１００回行い、１０回・２０回・５０回・１００回までの結果から表の出る確率をそれぞれ推定せよ。解答用紙には１００回分の結果と１０回・２０回・５０回・１００回のときの推定値を書くこと。

（表は「表」を「○」、裏を「×」でかまいません）

	1	2	3	4	5	6	7	8	9	10
0										
10										
20										
30										
40										
50										
60										
70										
80										
90										

表の出る確率 p の推定値（表を 1, 裏を 0 としたときの平均値）

・　１０回目まで

・　２０回目まで

・　５０回目まで

・１００回目まで

（分数ではなく有効桁数 3 桁の小数で答えること）