数理統計学　前回の解答

１）区間 $[2,6]$ の一様分布において、区間 $[2.2,5.3]$ に入る確率。

一様分布の確率密度関数は $f(x)=\dfrac{1}{6-2}=\dfrac{1}{4}$ $(2\leqq x\leqq 6)$ であるから

$$P(2.2\leqq X\leqq 5.3)=\int_{2.2}^{5.3}\frac{1}{4}\,dx=\left[\frac{x}{4}\right]_{2.2}^{5.3}=\frac{5.3-2.2}{4}$$
$$=\frac{3.1}{4}=\frac{31}{40}=0.775$$

２）パラメータ $\lambda=\frac{1}{3}$ の指数分布において、区間 $[3,9]$ に入る確率。

指数分布の確率密度関数は $f(x)=\frac{1}{3}e^{-\frac{1}{3}x}$ $(0\leqq x)$ であるから

$$P(3\leqq X\leqq 9)=\int_{3}^{9}\frac{1}{3}e^{-\frac{1}{3}x}\,dx=\left[-e^{-\frac{1}{3}x}\right]_{3}^{9}=-e^{-3}-\left(-e^{-1}\right)$$
$$=e^{-1}-e^{-3}=0.318092\cdots=0.318$$

注）解答は $e^{-1}-e^{-3}$ でも 0.318 でも可

３）標準正規分布 $N(0,1)$ において、区間 $[-1.68,-0.37]$ 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。

$$P(-1.68\leqq X\leqq -0.37)=P(0.37\leqq X\leqq 1.68) \quad \text{（分布は左右対称）}$$
$$=P(0\leqq X\leqq 1.68)-P(0\leqq X\leqq 0.37) \quad \text{（0までを足して引く）}$$
$$=0.45352-0.14431=0.30921=0.309 \quad \text{（表を使って確率の計算）}$$

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）。

$$P(-1.37\leqq X\leqq 0)=P(0\leqq X\leqq 1.37)=0.41466$$

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。

数理統計学 平均と分散　　第４回

・平均と分散の定義

　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 は次のように定義されている。

$$\mu = E[X] = \begin{cases} \sum x_i \times P(X = x_i) & \text{(離散型の場合)} \\ \int_{-\infty}^{\infty} x f(x)\, dx & \text{(離散型の場合)} \end{cases}$$

また、平均値からのバラツキ具合を表す分散 σ^2 は次のように定義されている。

$$\sigma^2 = E[(X-\mu)^2] = \begin{cases} \sum (x_i - \mu)^2 \times P(X = x_i) & \text{(離散型の場合)} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx & \text{(離散型の場合)} \end{cases}$$

なお分散に関しては、次のように計算することもできる　$\sigma^2 = E[X^2] - \mu^2$

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 $\frac{1}{6}$ で出現する離散型である。
よって、平均値 μ は

$$\mu = E[X] = \sum_{x=1}^{6} x \times P(X = x) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.50$$

また、分散の値は

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

$$= \left(1 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \cdots + \left(6 - \frac{7}{2}\right)^2 \times \frac{1}{6} = \frac{35}{12} = 2.9166\cdots = 2.92$$

例２．幾何分布の平均・分散

　確率は $P(X = k) = (1-p)^k p$ であることから、平均値 μ は

$$\mu = E[X] = \sum_{k=0}^{\infty} k \times P(X = k) = \sum_{k=0}^{\infty} k \times (1-p)^k p = \frac{1-p}{p}$$

さらに分散 σ^2 は

$$\sigma^2 = E[(X-\mu)^2] = E[X^2] - \mu^2 = \sum_{k=0}^{\infty} k^2 \times P(X = k) - \left(\frac{1-p}{p} \right)^2$$

$$= \sum_{k=0}^{\infty} k^2 \times (1-p)^k p - \frac{(1-p)^2}{p^2} = \frac{p^2 - 3p + 2}{p^2} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2}$$

例３．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

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

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

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。

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

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

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

一般に、母集団から計算された平均・分散を母平均・母分散と呼び、標本から計算された平均・分散を標本平均・標本分散と呼び、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）を標準偏差と呼ぶ。

《参考》　受験等でよく用いられる「偏差値」は

$$\frac{10(y-\bar{x})}{s}+50$$

によって計算されている。テストの点が正規分布に従っていると仮定した場合、偏差値 30〜70（つまり $\pm2\sigma$）までに約 95％の人が入っていて、偏差値 20〜80（つまり $\pm3\sigma$）までに 99％以上（正確には 99.73％）の人が入っている。したがって、ほとんどの場合偏差値は 20〜80 の間に収まるが、場合によっては偏差値が負になったり、100 以上になったりすることがある。

・Rを使って分布のグラフを書く

・連続型の場合

Rには基本的な分布の関数が既に用意されている。

一様分布：dunif(x, min=a, max=b)　パラメータを省略すると $(0,1)$ の一様分布

指数分布：dexp(x, λ)　パラメータを省略すると $\lambda = 1$

正規分布：dnorm(x, μ, σ)　パラメータを省略すると $N(0,1)$

注：パラメータは分散の σ^2 ではなく、標準偏差 σ を使っている）

Rにはグラフを書くコマンドとして "plot" と "curve" が用意されている。

コマンド："curve"　⇒　curve(関数, x の下限, x の上限)　主に連続型分布のグラフを描く

コマンド："plot"　⇒　plot(関数, x の下限, x の上限)　主に離散型分布のグラフを描く

注）plot は散布図なども描けるので、その場合は括弧の中の意味が変わる

例．一様分布

$(0,1)$ の一様分布の確率密度関数は

```
> curve(dunif, -1, 2)
```

$(2,5)$ の一様分布の確率密度関数は

```
> curve(dunif(x, 2, 5), 1, 6)
```

例．指数分布

$\lambda = 1$ の指数分布の確率密度関数は

```
> curve(dexp, 0, 5)
```

$\lambda = 0.7$ の指数分布の確率密度関数は

```
> curve(dexp(x, 0.7), 0, 5)
```

例．正規分布

標準正規分布 $N(0,1)$ の確率密度関数は

```
> curve(dnorm, -5, 5)
```

正規分布 $N(0,2^2)$ の確率密度関数は

```
> curve(dnorm(x, 0, 2), -10, 10)
```

《参考》一般の関数 $f(x) = x^2$ などは次のように関数を定義してから curve すればよい。

```
> f <- function(x) x^2      # 関数として x^2 を f と定義する
> curve(f ,-2, 2)
```

・離散型の場合Rには基本的な分布の関数が既に用意されている。

幾何分布 : dgeom(x, prob=p)　パラメータの省略不可

二項分布 : dbinom(x, n, prob=p)　パラメータの省略不可

コマンド："plot"

plot(x 軸のデータ, y 軸のデータ, オプション)

例. 幾何分布

$p = \frac{1}{2}$ の幾何分布の確率は

```
> plot(0:10,dgeom(0:10,prob=1/2),type="h",lwd=10,col="gray")
```

$p = \frac{1}{5}$ の幾何分布の確率は

```
> plot(0:10,dgeom(0:10,prob=1/5),type="h",lwd=10,col="gray")
```

例. 二項分布

$n = 5,\ p = \frac{1}{2}$ の二項分布の確率は

```
> plot(0:5,dbinom(0:5,5,prob=1/2),type="h",lwd=10,col="gray")
```

$n = 10,\ p = \frac{1}{4}$ の二項分布の確率は

```
> plot(0:10,dbinom(0:10,10,prob=1/4),type="h",lwd=10,col="gray")
```

グラフを重ねて書くときは 1 つめのグラフを書いた後に "par(new=T)" という上書きの指定をした後に 2 枚目のグラフを書く。ただし、x 軸と y 軸の範囲をあわせないと、範囲の違うグラフが重なってしまうので注意が必要（目盛りの部分が重なってしまう)。

悪い例

```
> curve(dnorm,-4,4)
> par(new=T)
> curve(dnorm(x,0,4),-12,12)
```

正しい例

```
> curve(dnorm,-12,12,ylim=c(0,0.5))
> par(new=T)
> curve(dnorm(x,0,4),-12,12,ylim=c(0,0.5))
```

今回の演習問題

正規分布 $N(0,1^2), N(0,4^2), N(3,2^2)$ の 3 つのグラフ x 軸の区間を $[-5,10]$ の範囲で 1 つのグラフに書きなさい。（どのグラフか分かるように説明を書くこと）

ヒント）y の範囲は ylim=c(0,0.5) もしくは ylim=c(0,0.6) あたりが良いです。

今回の演習は、グラフをＡ４用紙に印刷し、下の部分に学科・学籍番号・名前を記載して期限内に提出してください。（K3-3309 号室前　17 番のボックス）

Rで描いたグラフの印刷方法

１．直接印刷

・グラフの画面が選択されているときに、「ファイル」→「印刷」を選択する。

・グラフの画面で右クリックすると「印刷」が出てくるので、それを選択する。

ただし直接印刷では、グラフの大きさが選べない。

２．ワード等に貼り付けて印刷

1) グラフの画面で右クリックし、
「メタファイルにコピー」または「ビットマップにコピー」を選択

2) ワード等のソフトで「編集」→「貼り付け」を選択する。(貼り付けは別の方法でもＯＫ)

ワードの場合、「メタファイルにコピー」または「ビットマップにコピー」のどちらでも貼り付けることができるが、一部のソフトでは「ビットマップにコピー」しかできない場合がある。基本的には「メタファイルにコピー」がきれいに印刷できるので、両方試した方が良い。

今回の演習は、解答用紙に直接印刷しても良いし、別紙に印刷したものを貼り付けても良い(のり又はセロハンテープ)。ただし、機械で読み取って保存するためホッチキスの使用は不可とする。

演習提出用紙：

演習の際に配布した用紙（通常は色紙で配布）

もしくはＷｅｂページにある雛形を印刷したもの（コピー用紙に学年組名前を書いたものでも可）

Ａ４以外の用紙やルーズリーフは不可とする

演習提出先：

K3-3309 号室前　16 番のボックス（注：Ｎ／Ｄ科の数理統計学と共通）

演習提出期限：

基本的に月曜日１７時頃まで

注）演習の解説を次の授業時間に行うので、期限後の提出は減点対象となります。
ただし提出しない場合、その回の演習点は０点です。

2019年度 数理統計学 D科　小テスト解答用紙　2019.10.18

・Rを使って正規分布のグラフを描く

正規分布 $N(0,1^2), N(0,4^2), N(3,2^2)$ の３つのグラフ x 軸の区間を $[-5,10]$ の範囲で１つのグラフに書きなさい。（この紙でなく、Ａ４用紙に印刷したものに名前を書いて提出してもよい）

ヒント）資料をよく読んで 3 つの分布の違いがわかるグラフを描くこと。

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

提出先：K3-3309号室前　20番のボックス　提出期限：１０月２１日（月）１７時頃まで