数理統計学　前回の解答

第 5 回分

正規分布 $N(4, 2^2)$ に従う乱数を 10 個発生させて、その平均と分散を 10 回計算する。
さらに、計算された 10 回の平均・分散の平均と分散を計算する。

実際に乱数を発生させて、平均・分散を求める実験を 100,000 回行ったときのヒストグラムは次のようになる。（プログラムは裏面）

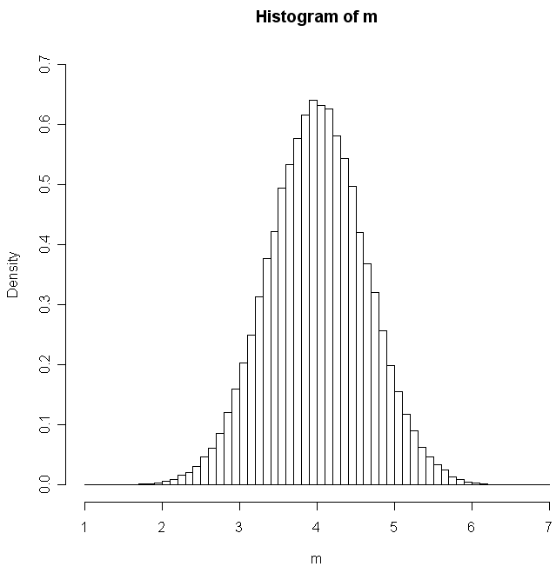

図 1: 平均のヒストグラム

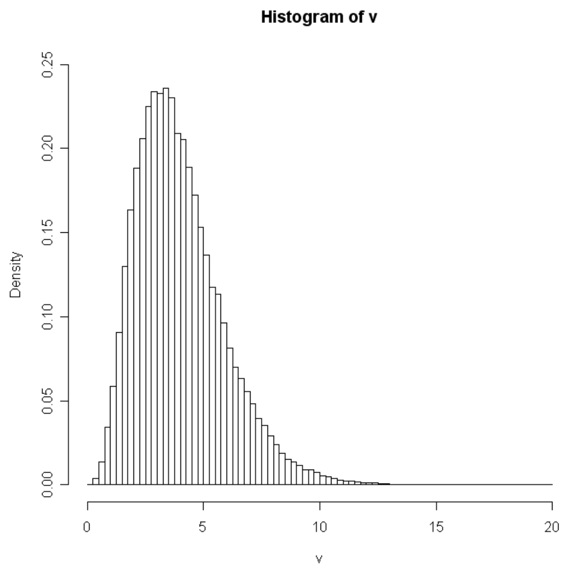

図 2: 分散のヒストグラム

正規分布 $N(\mu, \sigma^2)$ に従うとき、n 個のデータによる標本平均 $\bar{x}$ は正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従い、標本分散 s^2 は $\frac{n-1}{\sigma^2}s^2$ と変換すると自由度 $n-1$ の χ^2 分布に従うことが理論的にわかっている。今回の例題では母集団を $N(4, 2^2)$ としたので、標本平均 $\bar{x}$ は正規分布 $N(4, \frac{2^2}{10})$ に従っていることになる。よって、理論分布から精密な値を求めると

$[2.76, 5.24]$ の間に 95%の標本平均が入り、$[2.37, 5.63]$ の間に 99%の標本平均が入る。

また、標本分散 s^2 は

$[1.20, 8.45]$ の間に 95%の標本分散が入り、$[0.771, 10.5]$ の間に 99%の標本分散が入る。

ちなみに、

平均の平均 $\bar{\bar{x}}$ は

$[3.61, 4.39]$ の間に 95%が入り、$[3.48, 4.52]$ の間に 99%が入る。

平均の分散 $s^2_{\bar{x}}$ は

$[0.120, 0.845]$ の間に 95%が入り、$[0.0771, 1.05]$ の間に 99%が入る。

また、シミュレーションによって

分散の平均 $\bar{s^2}$ は

$[2.92, 5.26]$ の間に 95%が入り、$[2.62, 5.72]$ の間に 99%が入る。

分散の分散 $s^2_{s^2}$ は

$[0.889, 8.98]$ の間に 95%が入り、$[0.566, 12.4]$ の間に 99%が入る。

・ヒストグラムを書くプログラム

配列 m と v を用意して、その中に 10 個のデータの平均と分散を代入することを 10 万回繰り返す。

```
m <- numeric(100000)
v <- numeric(100000)
for(i in 1:100000) {
  x <- rnorm(10,4,2)
  m[i] <- mean(x)
  v[i] <- var(x)      }
```

標本平均のヒストグラムを書く（0 から 8 まで 0.2 刻み）

```
hist(m,breaks = seq(0,8,0.2),freq=F,ylim=c(0,0.7))
```

標本分散のヒストグラムを書く（0 から 45 まで 0.5 刻み）

```
hist(v,breaks = seq(0,45,0.5),freq=F,ylim=c(0,0.12))
```

範囲外のデータがある場合、エラーとなるので、seq(a,b,c) の a,b の値を変えて区間の幅を少し広くするとよい。

さらに、グラフの重ねあわせをするのであれば、

・標本平均 $\bar{x}$ は $N(4, \frac{2^2}{10})$ のグラフを重ねればよいので

```
hist(m,breaks = seq(0,8,0.2),freq=F,ylim=c(0,0.7))
par(new=T)
curve(dnorm(x,4,sqrt(2^2/10)),0,8,ylim=c(0,0.7),ylab="")
```

・標本分散 s^2 は $\frac{9}{4}s^2$ で変換して自由度 9 の χ^2 分布のグラフを重ねればよいので

```
vv <- 9/2^2*v
hist(vv,breaks = seq(0,50,0.5),freq=F,ylim=c(0,0.12))
par(new=T)
curve(dchisq(x,9),0,50,ylim=c(0,0.12),ylab="")
```

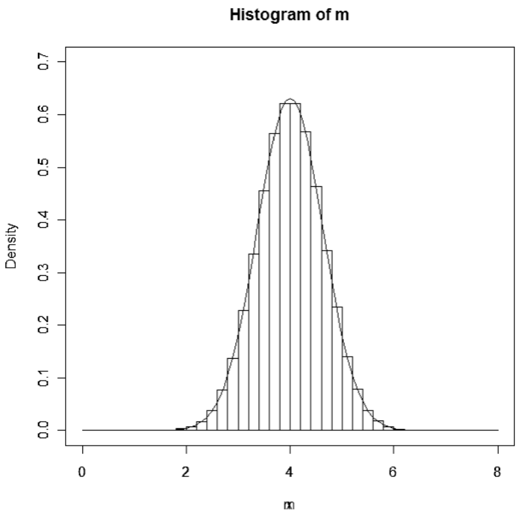

図 3: 平均に関するグラフ

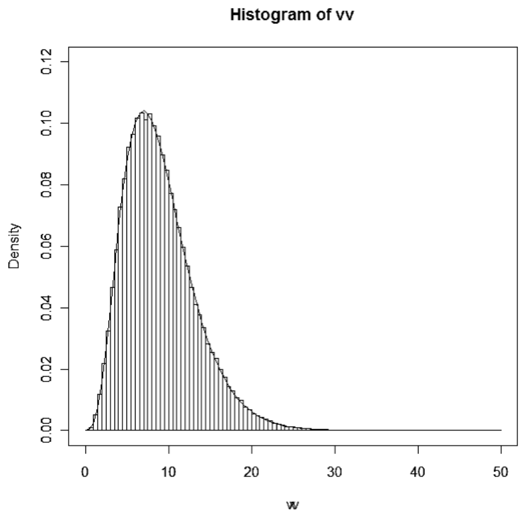

図 4: 分散に関するグラフ

学生番号別の解答

学籍番号	1 回目の平均	10 回目の分散	平均の平均	平均の分散	分散の平均	分散の分散
1822008	4.79	4.82	4.42	0.269	4.04	4.41
1822015	5.11	1.91	4.21	0.301	3.81	3.06
1822024	2.87	5.98	4.21	0.633	4.23	5.82
1822031	4.26	5.23	4.15	0.371	4.09	3.83
1822050	3.49	4.25	3.83	0.142	3.76	2.25
1822065	4.25	6.18	4.10	0.136	3.76	2.64
1822074	4.15	3.14	4.10	0.365	3.98	3.33
1822078	4.31	3.38	3.81	0.148	4.00	2.13
1822094	5.08	5.68	4.27	0.615	4.24	3.63
1822096	2.92	2.62	3.62	0.343	3.36	0.686
1822097	3.99	6.98	3.98	0.813	3.08	2.99
1822100	3.35	3.13	4.00	0.685	4.05	3.80

注）平均と分散の値を求めるときの四捨五入に **R** の関数を使用しています。
そのため、結果が微妙に異なる可能性があります。

・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 $x_1, x_2, \cdots, x_n$ が得られたときの尤度関数は

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

であり、対数尤度関数は

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

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

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

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

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

ちなみに、指数分布の母平均は $\dfrac{1}{\lambda}$ 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》

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

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。

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

数理統計学　正規分布の確率と平均値の推定

・正規分布の確率計算

正規分布の確率を教科書 P.172 にある標準正規分布の確率を使って求めることができる。この授業においては、R を使うことによって、もっと簡単に近似値を求めることができる。

正規分布の確率を求める R の関数は $N(\mu,\sigma^2)$ に従うときの分布関数、すなわち $F(x) = P(-\infty < X \leqq x)$ の値を `pnorm(x,`μ`,`σ`)` で求められる。

実際に $N(\mu,\sigma^2)$ に従う場合、$P(a < X \leqq b)$ の値は
$P(a < X \leqq b) = P(-\infty < X \leqq b) - P(-\infty < X \leqq a)$ なので

```
> pnorm(b,μ,σ)-pnorm(a,μ,σ)
```

によって、求めることができる。

例題．教科書にある通り、１８才女性の身長が $N(157.0,(5.0)^2)$ の正規分布に従うとき

１）155cm 以上 170cm 以下である確率を求めよ

$P(155 \leqq X \leqq 170)$ を求めればよいので、

```
> pnorm(170,157,5)-pnorm(155,157,5)
 [1] 0.6507606
```

のように $0.651 = 65.1\%$ であることが分かる。

２）160cm 以下である確率を求めよ

$P(X \leqq 160)$ を求めればよいので、

```
> pnorm(160,157,5)
 [1] 0.7257469
```

のように $0.726 = 72.6\%$ であることが分かる。

・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

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

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、$[\bar{x} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}]$ 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。

・R を使った区間推定の求め方

基本的に必要な計算は正規分布や t 分布の確率計算である。R には基本的な分布の確率が与えられた場合の区間を簡単に計算できる。

・$P(-z_\alpha < Z < z_\alpha) = 1-\alpha$ となる z_α の求め方

R では $P(-\infty < Z < z_\beta) = \beta$ となる z_β を `qnorm(`β`)` で求めることができる。$P(-z_\alpha < Z < z_\alpha) = 1-\alpha$ となる z_α を求めるときは、正規分布の確率密度関数が左右対称である事を使って、

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

で求められる。

例. 標準正規分布 $N(0,1)$ で $P(-z_\alpha < Z < z_\alpha) = 0.95 = 1-0.05$ となる z_α は $1-0.05/2 = 0.975$ であるから、`qnorm(0.975)` で求めることができる。

```
> qnorm(0.975)
[1] 1.959964
```

例１．母平均が既知の場合の区間推定

母集団が母分散 3^2 の正規分布に従うとする。そこから無作為に 10 個のデータが得られ、標本平均が 17.44 だったとする。このとき、母平均 μ の 99%信頼区間を求めよ。

$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}-\mu}{\frac{3}{\sqrt{10}}}$$

が標準正規分布に従う。よって、$P(-z_\alpha < Z < z_\alpha) = 0.99 = 1-0.01$ となる z_α は

$$P(Z < z_\alpha) = 1 - \frac{0.01}{2} = 0.995 \quad \Rightarrow \texttt{qnorm(0.995)} = 2.575829$$

となる。よって母平均の 99%信頼区間は

$$[17.44 - 2.575829 \times \frac{3}{\sqrt{10}}, 17.44 + 2.575829 \times \frac{3}{\sqrt{10}}] \quad \Rightarrow \quad [14.99635, 19.88365]$$

有効数字 3 桁で答えると $[15.0, 19.9]$ である。

資料置場

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

2019年度 数理統計学 N科　小テスト解答用紙　2019.10.30

・正規分布の確率と平均の区間推定

１）身長の分布が $N(170.5, 7.5^2)$【男性の場合】、$N(156.5, 6.5^2)$【女性の場合】、に従うと仮定できる場合、自分の身長《○○○ cm（小数第１位を四捨五入)》以上になる確率 $P(○○○ \leqq X)$ を求めよ

２）母集団が母分散 $(3.25)^2$ の正規分布に従うとき、９つのデータ

32.8, 41.5, 32.5, 38.3, 37.0, 40.6, 36.9, 33.1, 36.7

が与えられたときの、標本平均を計算し、母平均 μ の 95%信頼区間を求めよ。

注）どちらの解答もどのような式もしくはプログラムを使ったか簡単に記載すること。

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

提出先：K3-3309号室前　19番のボックス　提出期限：１１月　５日（火）１７時頃まで