確率統計と情報処理・演習（2007 年度後期）

確率分布

2007 年 12 月 21 日

日本女子大学理学部数物科学科　　今野 良彦

November 30, 2007

今日の講義の目的と概要

- 推定について
  - 母集団と標本
  - 母数（母平均・母分散）
  - 推定量と推定値
  - 点推定量：性質と推定量の導出について
    - 不偏推定量
    - 有効推定量：最良線形不偏推定量（ここまでは先週）
  - 区間推定（ここからは今週）
    - 信頼係数と信頼区間
  - 正規母集団に対する母平均の区間推定
    - 母分散が既知の場合の母平均の区間推定
    - 母分散が未知の場合の母平均の区間推定

先週の内容の復習：母集団と標本

⋆ 母集団：ある集団の特徴を調査したいとき，興味の対象の集団の全体

⋆ 例：

- 2006 年度に日本全体で小学校に入学したときの生徒の身長の平均が知りたい ⟶ 2006 年度に日本全体で小学校に入学したときの生徒全員が母集団
- ある工場で生産されている電球の寿命が知りたい！たとえば，1000 時間以上の寿命があるかどうか？ ⟶ 母集団は母集団は生産される電球すべて！

⋆ 全数調査：母集団の全員を調査すること

⋆ 標本調査：母集団から抜き出された一部分だけを調査すること

⋆ 標本：母集団から抜き出された一部分のこと

先週の内容の復習：推定とは

⋆　母集団の個体のすべてを調査しない！

⋆　母集団の個体の特性値の分布を確率分布でモデル化（まねる）する．この確率分布のことを**母集団分布**という！

⋆注意： 分布とは，個体の特性値がどのあたりにどのくらいあるかを表現するものである！

⋆ 目標：母集団分布を特定したい！できなければ，母集団分布の平均や分散を推測したい！

⋆ 母数：母集団を特定するために役にたつ指標のことをいう．例としては，母集団分布の平均と分散．母集団分布の確率密度関数も母数の例となる．特にことわらないかぎり，推定の対象となる母数は 1 次元として議論を進めていく．

⋆ 母平均：母集団分布の平均

⋆ 母分散：母集団分布の分散

先週の内容の復習：推定法について

- 点推定法 — 標本に基づいてひとつの値で母数の値を推定
- 区間推定法 —- 推定したい母数が「ある確率で入っている」区間を求めること．たとえていえば、多く見積もればいくつ，小さく見積もればいくつといった具合

区間推定法とは

⋆ 点推定法では，一組のデータから 1 つの値を定めることで母数を推定していた．だった一組のデータで求めた値は母数に一致する可能性は稀である．

⋆ 母集団の推定したい母数がある確率で入っている区間を構成する「区間推定」を考えてみよう．

⋆ 理想的には，定めた区間に母数が入っている確率をできるだけ大きくして，区間の幅を短くできるとよい．

しかし，

⋆ この両方の希望を同時にみたすことはできない！

⋆ すなわち，区間の幅を狭くすると，区間が母数を含んでいる確率は下がり，逆に，区間が母数を含んでいる確率をあげると区間の幅は長くなる．

⋆ では，どのように区間をさだめるのだろうか？

区間推定の考え方

⋆ 信頼区間のなかに母数の入っている確率に条件をつけることにする．

⋆ 信頼度（信頼係数）— 信頼区間に母数が入っている確率が $1-\alpha$ 以上になるように要請する精度保障の基準．ただし，$0<\alpha<1$.

⋆ 実用上は，信頼係数は $0.95\,(\alpha=0.05)$ または $0.99\,(\alpha=0.01)$ を用いることがおおい．

⋆ n 個の標本 $X_1, X_2, \dots, X_n\,(n\ge 2)$ に基づいて，母数 θ が入っていると思われる区間を

$$(\widehat{\theta}_L, \widehat{\theta}_U)$$

とする．ただし，$\widehat{\theta}_L, \widehat{\theta}_U$ の値域は 1 次元の統計量（確率変数）で

$$\widehat{\theta}_L := \widehat{\theta}_L(X_1, X_2, \dots, X_n), \qquad \widehat{\theta}_U := \widehat{\theta}_U(X_1, X_2, \dots, X_n).$$

信頼係数の条件：信頼度 $1-\alpha$ の信頼区間

$$\mathbb{P}(\widehat{\theta}_L < \theta < \widehat{\theta}_U) \ge 1-\alpha \tag{1}$$

⋆ 条件 (1) を満たす信頼区間（信頼度 $1-\alpha$ の信頼区間）の中で，区間の幅

$$\widehat{\theta}_U - \widehat{\theta}_L$$

をできるだけ短くした区間 $(\widehat{\theta}_L, \widehat{\theta}_U)$ を定めたい！

ここまでのまとめ

⋆ 信頼度（信頼係数）はわかりましたか？

⋆ 信頼区間がみたす条件は理解しましたか？

⋆ 仮定：個体の特性値は「連続量」で，母集団分布は正規分布と仮定する．

この仮定より

母集団分布の母平均と母分散がわかる $\Longleftrightarrow$ 母集団分布がわかる！

なぜならば，正規分布の確率密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, \qquad -\infty < x < \infty$$

で

$$\begin{aligned} \text{母平均：}\quad \mu &= \int_{-\infty}^{\infty} x f(x)\,dx \\ \text{母分散：}\quad \sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \end{aligned}$$

母平均 μ の信頼区間：母分散 σ^2 が既知の場合

⋆ 標本 $X_1, X_2, \dots, X_n\,(n\ge 2)$ が平均が μ，分散が σ^2 の正規分布 $N(\mu, \sigma^2)$ から独立にとられた標本とする．

⋆ すなわち，任意の $a_i, b_i \in \mathbb{R}\,(i=1,2,\dots,n)$ に対して，

$$\begin{aligned} &\mathbb{P}(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, \dots, a_n < X_n \le b_n) \\ &\quad = \Pi_{i=1}^n \mathbb{P}(a_i < X_i \le b_i) \\ &\quad = \Pi_{i=1}^n \int_{a_i}^{b_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \end{aligned}$$

⋆ このとき，標本平均

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

は

$$N\left(\mu, \frac{\sigma^2}{n}\right)$$

に従う．

なぜならば，

⋆ 正規分布に従う確率変数の線形結合により作られた確率変数の分布は正規分布．

⋆ 正規分布は平均と分散がわかれば，分布が特定できる！

⋆ $\mathbb{E}[\bar{X}_n] = \mathbb{E}[(1/n)\sum_{i=1}^n X_i] = (1/n)\sum_{i=1}^n \mathbb{E}[X_i] = \mu$
と
$\mathbb{VAR}[\bar{X}_n] = \mathbb{VAR}[(1/n)\sum_{i=1}^n X_i] = (1/n)^2 \sum_{i=1}^n \mathbb{VAR}[X_i] = \sigma^2/n$

これより，

$$Z=\frac{\bar{X}_n-\mathbb{E}[\bar{X}_n]}{\sqrt{\mathbb{VAR}[\bar{X}_n]}}=\frac{\bar{X}_n-\mu}{\sqrt{\sigma^2/n}}\sim N(0,\,1^2)$$

⋆ なぜならば，Z の確率密度関数を $f_Z(z)$ とする．$-\infty<z<\infty$ に対し，

$$f_Z(z)=\frac{d}{dz}\mathbb{P}(Z\le z)$$

に注意.

$$\begin{aligned}\mathbb{P}(Z\le z) &= \mathbb{P}\left(\frac{\bar{X}_n-\mu}{\sqrt{\sigma^2/n}}\le z\right)=\mathbb{P}\left(\bar{X}_n\le\sqrt{\sigma^2/n}z+\mu\right)\\ &= \int_{-\infty}^{\sqrt{\sigma^2/n}z+\mu}\frac{1}{\sqrt{\sigma^2/n}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2/n}\right\}dx\end{aligned}$$

よって

$$f_Z(z)=\frac{d}{dz}\int_{-\infty}^{\sqrt{\sigma^2/n}z+\mu}\frac{1}{\sqrt{2\pi(\sigma^2/n)}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2/n}\right\}dx=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}$$

確率変数 Z が標準正規分布 $N(0,\,1^2)$ に従うとき，

$$\mathbb{P}(a<Z<b)=1-\alpha$$

をみたす数 $a,\,b$ は無数にある．

⋆ しかし，$b-a$ を最小にするには，山の高いところを中心に

$$\mathbb{P}(-k<Z<k)=1-\alpha$$

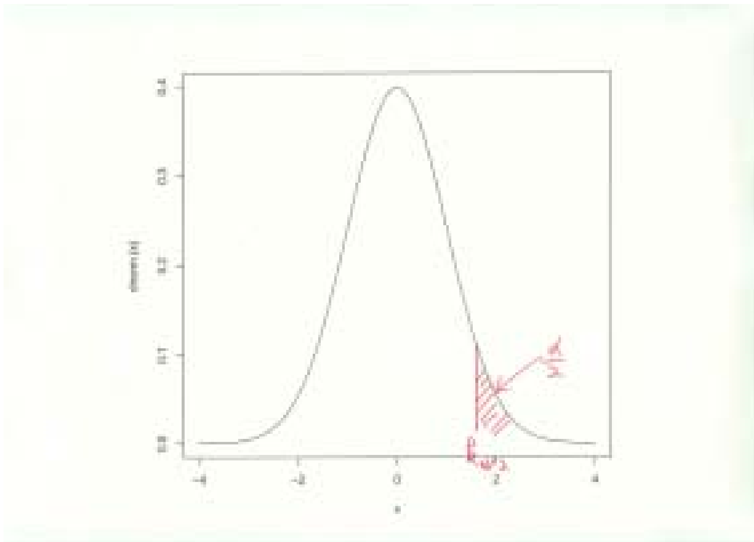
ととればよい！

	点	確率 $\mathbb{P}(Z\le z)$	確率 $\mathbb{P}(Z\le z)$	
上限	1.959964	0.975	2.326348	0.99
下限	-1.959964	0.025	-1.750686	0.04
	3.919928	0.95	4.077034	0.95

⋆ $0<\alpha<1$ に対し

$$\mathbb{P}(Z>k_{\alpha/2})=\frac{\alpha}{2}$$

なる上側 $\frac{100\alpha}{2}$ % 点をもとめればよい．



⋆ $0<\alpha<1$ に対し

$$\mathbb{P}(Z>k_{\alpha/2})=\frac{\alpha}{2}$$

なる上側 $\frac{100\alpha}{2}$% 点を用いて式の変形をする：

$$\begin{aligned}1-\alpha &= 1-\frac{\alpha}{2}-\frac{\alpha}{2}=1-\mathbb{P}(Z>k_{\alpha/2})-\mathbb{P}(Z>k_{\alpha/2})\\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt-\int_{k_{\alpha/2}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt-\int_{k_{\alpha/2}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt\\ &= \int_{-\infty}^{k_{\alpha/2}}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt-\int_{-\infty}^{-k_{\alpha/2}}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt\\ &= \int_{-k_{\alpha/2}}^{k_{\alpha/2}}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt=\mathbb{P}(-k_{\alpha/2}<Z<k_{\alpha/2})\end{aligned}$$

つぎに，

$$1-\alpha=\mathbb{P}(-k_{\alpha/2}<Z<k_{\alpha/2})$$

に $Z=(\bar{X}_n-\mu)/\sqrt{\sigma^2/n}$ を代入すると

$$\begin{aligned}1-\alpha &= \mathbb{P}\left(-k_{\alpha/2}<\frac{\bar{X}_n-\mu}{\sqrt{\sigma^2/n}}<k_{\alpha/2}\right)\\ &= \mathbb{P}\left(-k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}<\bar{X}_n-\mu<k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)\\ &= \mathbb{P}\left(-\bar{X}_n-k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}<-\mu<-\bar{X}_n+k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)\\ &= \mathbb{P}\left(\bar{X}_n+k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}>\mu>\bar{X}_n-k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)\end{aligned}$$

すなわち

$$1-\alpha = \mathbb{P}\left(\bar{X}_n - k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X}_n + k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$

この式は，推定したい母平均 μ が区間

$$\left(\bar{X}_n - k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}},\ \bar{X}_n + k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$

に確率 $1-\alpha$ ではいっていることを示している！

信頼係数 $(1-\alpha)\times 100\%$ の母平均 μ の信頼区間

$$\left(\bar{X}_n - k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}},\ \bar{X}_n + k_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$

母分散が既知のときの母平均の信頼区間のシミュレーション

⋆ 標本分布を $N(0,1)$ としたとき，標本数 $n=10$ の信頼係数 0.95 の母平均の信頼区間

信頼区間

```
> k<-qnorm(1-0.025,0,1)
> k
[1] 1.959964
> x<-rnorm(10,0,1)
> x
 [1] -0.95275764  0.39604400  2.15215811 -0.22070032  0.22461379
 [6] 0.06671838  -2.86975585  0.92568507 -2.54089931 -0.98980498
> barx<-mean(x)
> barx
[1] -0.3808699
> mL<-barx-k*sqrt(1/10)
> mU<-barx+k*sqrt(1/10)
> mL      # 信頼区間の下側限界
[1] -1.000665
> mU      # 信頼区間の上側限界
[1] 0.2389252
```


区間

$$(-1.000665,\ 0.2389252)$$

は母平均 μ の信頼度 95% の信頼区間となる．

標本 $(x_1, x_2, \ldots, x_n)$ と信頼度 $(1-\alpha)100\%$ の値 $1-\alpha = \mathrm{conf}$ と既知の母分散 σ^2 の値 sigma2 を与えて，信頼区間を求める関数

conf.interval

```
> conf.interval<-function(x,conf,sigma2){
+ n<-length(x)
+ alpha<-1-conf
+ k<-qnorm(1-alpha/2,0,1)
+ barx<-mean(x)
+ mL<-barx-k*sqrt(sigma2/n)
+ mU<-barx+k*sqrt(sigma2/n)
+ return(c(mL,mU))
+ }
> conf.interval(x,0.95,1)
[1] -1.0006649  0.2389252
>
```


conf.interval の実行

```
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -0.4407088  0.7988812
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -0.7284778  0.5111122
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -0.7529757  0.4866144
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -1.0780105  0.1615796
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -1.28314464 -0.04355458
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -0.6162157  0.6233744
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -1.0766413  0.1629487
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -0.9572526  0.2823375
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -0.6586764  0.5809137
> conf.interval(rnorm(10,0,1),0.95,1)
[1] -0.4116793  0.8279107
```


信頼係数と信頼区間の幅の関係をみてみよ．

conf.interval の実行

```
> x
 [1] -0.95275764  0.39604400  2.15215811 -0.22070032  0.22461379
 [6] 0.06671838  -2.86975585  0.92568507 -2.54089931 -0.98980498
 > y080<-conf.interval(x,0.80,1)
> y080
[1] -0.78613206  0.02439231
> y080[2]-y080[1]
[1] 0.8105244
> y090<-conf.interval(x,0.90,1)
> y090
[1] -0.9010183  0.1392785
> y090[2]-y090[1]
[1] 1.040297
> y095<-conf.interval(x,0.95,1)
> y095
[1] -1.0006649  0.2389252
> y095[2]-y095[1]
[1] 1.23959
> y099<-conf.interval(x,0.99,1)
> y099
[1] -1.1954186  0.4336789
> y099[2]-y099[1]
[1] 1.629097
```


$N(0,1^2)$ から標本数 $n=10$ の標本を 100 回繰り返し抽出し，信頼度95% の信頼区間を 100 個作成する．

conf.interval

```
> for (i in 1:100){
+ print(conf.interval(rnorm(10,0,1),0.95,1))
+ }
```



```
[1] -0.4103589  0.8292311
[1] -0.2691124  0.9704777
[1] -0.6794240  0.5601661
[1] -1.1243018  0.1152883
[1] -1.14312049  0.09646958
[1] -0.6052396  0.6343504
[1] -0.4836565  0.7559336
[1] -0.6724574  0.5671326
[1] -1.15010483  0.08948523
[1] -0.4561505  0.7834396
[1] -0.3101039  0.9294861
[1] -0.6458842  0.5937059
[1] -0.2816609  0.9579292
[1] -0.5263421  0.7132480
[1] -0.6323301  0.6072600
[1] -0.5140249  0.7255651
[1] -0.6630502  0.5765399
[1] -1.30118369 -0.06159363       <- 母平均を含まない区間
[1] -0.7856636  0.4539265
[1] -0.4913147  0.7482753
[1] -0.7176052  0.5219849
[1] -0.7251859  0.5144042
```



```
[1] -0.4917693  0.7478208
[1] -0.1759157  1.0636743
[1] -0.664813  0.574777
[1] -0.5545437  0.6850463
[1] -0.8383094  0.4012807
[1] -0.5151586  0.7244315
[1] -0.5340719  0.7055181
[1] -0.1319127  1.1076774
[1] -0.9674491  0.2721410
[1] -0.3700513  0.8695388
[1] -0.6378694  0.6017206
[1] -1.17877693  0.06081314
[1] -0.7522586  0.4873315
[1] -0.8342115  0.4053785
[1] -0.7699872  0.4696029
[1] -0.8774342  0.3621558
[1] 0.01801523 1.25760530       <- 母平均を含まない区間
[1] -0.4779354  0.7616547
[1] -0.5639029  0.6756872
[1] -1.0979876  0.1416024
[1] -0.2596586  0.9799315
[1] -0.07908651  1.16050355
```



```
[1] -0.6088526  0.6307374
[1] -0.5938790  0.6457111
[1] -1.1309118  0.1086782
[1] -1.0618950  0.1776950
[1] -0.09044292  1.14914714
[1] -1.22267372  0.01691634
[1] -0.764919  0.474671
[1] -0.1976989  1.0418912
[1] -0.09907568  1.14051439
[1] -0.5158186  0.7237714
[1] -0.740364  0.499226
[1] -0.9535374  0.2860526
[1] -1.0523093  0.1872807
[1] -0.7154613  0.5241288
[1] -0.8744612  0.3651289
[1] -0.4538198  0.7857703
[1] 0.2483399 1.4879300       <- 母平均を含まない区間
[1] -0.8235227  0.4160674
[1] -0.2575735  0.9820166
[1] -0.5348435  0.7047466
[1] -1.1125506  0.1270395
[1] -0.8320479  0.4075421
```



```
[1] -0.1435447  1.0960453
[1] -1.14772481  0.09186526
[1] -0.9584142  0.2811758
[1] -1.0835302  0.1560599
[1] -0.4928898  0.7467002
[1] -0.9616206  0.2779694
[1] -0.8527472  0.3868428
[1] -0.5195266  0.7200635
[1] 0.09238219 1.33197225       <- 母平均を含まない区間
[1] -0.4542385  0.7853516
[1] -0.2368548  1.0027353
[1] -0.7633375  0.4762525
[1] -0.3599203  0.8796698
[1] -0.5056557  0.7339343
[1] -0.9296644  0.3099256
[1] -0.9254462  0.3141438
[1] -0.3374106  0.9021794
[1] -0.6379152  0.6016749
[1] -0.7327016  0.5068884
[1] -1.29052943 -0.05093936       <- 母平均を含まない区間
[1] -0.4382463  0.8013437
[1] -0.8462166  0.3933735
```

```
  [1] -0.9631503  0.2764397
  [1] -0.4868351  0.7527550
  [1] -0.3243519  0.9152382
  [1] -0.4026134  0.8369767
  [1] -0.4668755  0.7727146
  [1] -0.886738  0.352852
  [1] -0.8810452  0.3585449
  [1] -0.4334746  0.8061154
  [1] -0.9771255  0.2624646
  [1] -1.32157109 -0.08198103      <- 母平均を含まない区間
  [1] -0.3524163  0.8871738
  [1] -0.9724289  0.2671612
  >
```

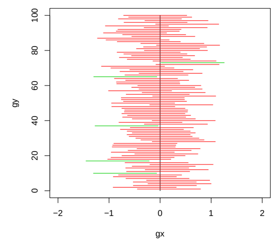



Figure 1: $N(0, 1)$ から $n = 10$ の標本に基づく信頼係数 95% の信頼区間の 100 個のグラフ

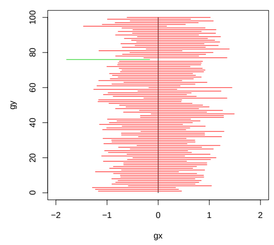


Figure 2: $N(0, 1)$ から $n = 10$ の標本に基づく信頼係数 99% の信頼区間の 100 個のグラフ

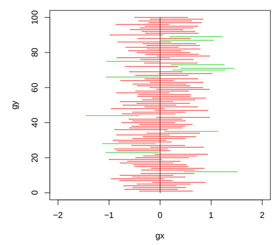


Figure 3: $N(0, 1)$ から $n = 10$ の標本に基づく信頼係数 80% の信頼区間の 100 個のグラフ

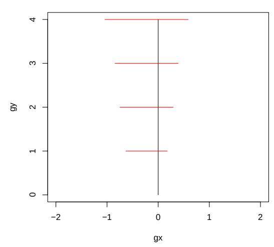


Figure 4: $N(0, 1)$ から $n = 10$ の標本に基づく信頼係数 99%, 95%, 90%, 80% の信頼区間(上から順番に)の 100 個のグラフ

ここまでのまとめ

⋆ 信頼度（信頼係数）はわかりましたか？

⋆ 信頼区間がみたす条件は理解しましたか？

母集団分布が正規分布で母分散が既知のときにどのような考え方で信頼区間が構成されるかがわかりましたか？

⋆ 信頼区間がある確率で母数を含むということの意味がわかりましたか？

⋆ 信頼係数と信頼区間の幅の関係がわかりましたか？

母平均 μ の信頼区間：母分散 σ^2 が未知の場合

$X_1, X_2, \ldots, X_n\,(n \geq 2)$ が $N(\mu, \sigma^2)$ に独立同一に従っている．

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

とする．母分散 σ^2 が既知の場合に，

$$Z = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,\, 1^2)$$

を利用した．

母分散 σ^2 が未知のときには，

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

の σ^2 のところに

$$\widehat{\sigma}_U^2 = \frac{1}{n-1}\sum_{i=1}(X_i - \bar{X}_n)^2$$

を代入する．

$$T := \frac{\bar{X}_n - \mu}{\sqrt{\frac{\widehat{\sigma}_U^2}{n}}}$$

すると

$$\begin{aligned}
T &= \frac{\bar{X}_n - \mu}{\sqrt{\frac{\widehat{\sigma}_U^2}{n}}} = \frac{(\bar{X}_n - \mu)\Big/\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{1}{n-1}\sum_{i=1}(X_i - \bar{X}_n)^2}}, \\
U &:= (\bar{X}_n - \mu)\Big/\sqrt{\frac{\sigma^2}{n}} \sim N(0,\, 1^2), \\
V &:= \frac{1}{n-1}\sum_{i=1}(X_i - \bar{X}_n)^2 \quad \text{は自由度 } (n-1) \text{ のカイ自乗分布} \\
& \quad\ U \text{ と } V \text{ は独立}
\end{aligned}$$

よって，$T = \dfrac{\bar{X}_n - \mu}{\sqrt{\frac{\widehat{\sigma}_U^2}{n}}}$　は自由度 $(n-1)$ の t 分布に従う

$t_{n-1}(\alpha/2)$: 自由度 $(n-1)$ の t 分布の上側点：すなわち，$0 < \alpha < 1$ に対し，

$$\mathbb{P}(T > t_{n-1}(\alpha/2)) = \frac{\alpha}{2}$$

自由度 $(n-1)$ の t 分布に従う確率変数 T の確率密度関数を $f_{T,\,n-1}(t)$ と書くことにする．

すると

$$\mathbb{P}(T > t_{n-1}(\alpha/2)) = \int_{t_{n-1}(\alpha/2)}^{\infty} f_{T,\,n-1}(t)\,dt$$

と $f_{T,\,n-1}(t) = f_{T,\,n-1}(-t)\;({}^\forall t \in \mathbb{R})$ に注意すれば，

$$\begin{aligned}
&\mathbb{P}(-t_{n-1}(\alpha/2) < T < t_{n-1}(\alpha/2)) \\
&\quad = \int_{-t_{n-1}(\alpha/2)}^{t_{n-1}(\alpha/2)} f_{T,\,n-1}(t)\,dt \\
&\quad = \int_{-\infty}^{\infty} f_{T,\,n-1}(t)\,dt - \int_{-\infty}^{-t_{n-1}(\alpha/2)} f_{T,\,n-1}(t)\,dt - \int_{t_{n-1}(\alpha/2)}^{\infty} f_{T,\,n-1}(t)\,dt \\
&\quad = \int_{-\infty}^{\infty} f_{T,\,n-1}(t)\,dt - \int_{t_{n-1}(\alpha/2)}^{\infty} f_{T,\,n-1}(t)\,dt - \int_{t_{n-1}(\alpha/2)}^{\infty} f_{T,\,n-1}(t)\,dt \\
&\quad = 1 - \alpha
\end{aligned}$$

$0 < \alpha < 1$ に対し，

$$\mathbb{P}(T > t_{n-1}(\alpha/2)) = \frac{\alpha}{2}$$

としたとき，

$$\mathbb{P}(-t_{n-1}(\alpha/2) < T < t_{n-1}(\alpha/2)) = 1 - \alpha$$

つぎに

$$T = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\widehat{\sigma}_U^2}{n}}}$$

を代入して変形していく．

$$
\begin{aligned}
& -t_{n-1}(\alpha/2) < T < t_{n-1}(\alpha/2) \\
& \quad \Longleftrightarrow -t_{n-1}(\alpha/2) < \frac{\bar{X}_n - \mu}{\sqrt{\frac{\widehat{\sigma}_U^2}{n}}} < t_{n-1}(\alpha/2) \\
& \quad \Longleftrightarrow -t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}} < \bar{X}_n - \mu < t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}} \\
& \quad \Longleftrightarrow -\bar{X}_n - t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}} < -\mu < -\bar{X}_n + t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}} \\
& \quad \Longleftrightarrow \bar{X}_n + t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}} > \mu > \bar{X}_n - t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}}
\end{aligned}
$$

よって，

$$
\begin{aligned}
1-\alpha &= \mathbb{P}(-t_{n-1}(\alpha/2) < T < t_{n-1}(\alpha/2)) \\
&= \mathbb{P}\left(\bar{X}_n - t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}} < \mu < \bar{X}_n + t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}}\right)
\end{aligned}
$$

母分散が未知のときの母平均 μ の信頼係数 $1-\alpha$ の信頼区間

$t_{n-1}(\alpha/2)$: 自由度 $(n-1)$ の t 分布の上側点：すなわち，$0 < \alpha < 1$ に対し，

$$
\mathbb{P}(T > t_{n-1}(\alpha/2)) = \frac{\alpha}{2},\ \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \widehat{\sigma}_U^2 = \frac{1}{n-1}\sum_{i=1}(X_i - \bar{X}_n)^2
$$

このとき，

$$
\left(\bar{X}_n - t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}},\ \bar{X}_n + t_{n-1}(\alpha/2)\sqrt{\frac{\widehat{\sigma}_U^2}{n}}\right)
$$

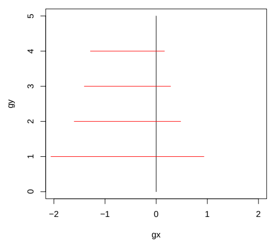


Figure 5: $N(0,1)$ から $n=10$ の標本に基づく信頼係数 99%, 95%, 90%, 85% の信頼区間(下から順番に)の 4 個のグラフ

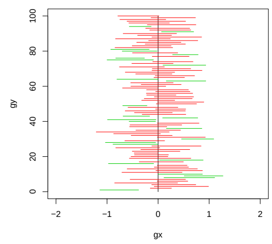


Figure 6: $N(0,2)$ から $n=10$ の標本に基づく信頼係数 80% の信頼区間の 100 個のグラフ

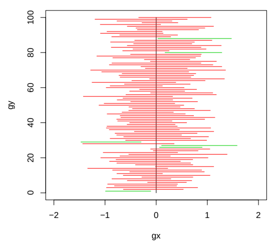


Figure 7: $N(0,2)$ から $n=10$ の標本に基づく信頼係数 95% の信頼区間の 100 個のグラフ

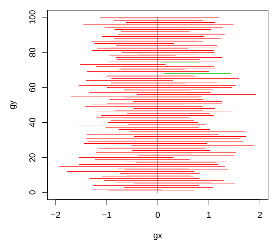


Figure 8: $N(0,2)$ から $n=10$ の標本に基づく信頼係数 99% の信頼区間の 100 個のグラフ

ここまでのまとめ

⋆ 信頼度（信頼係数）はわかりましたか？

⋆ 信頼区間がみたす条件は理解しましたか？

母集団分布が正規分布で母分散が未知のときにどのような考え方で信頼区間が構成されるかがわかりましたか？

⋆ 信頼区間がある確率で母数を含むということの意味がわかりましたか？

⋆ 信頼係数と信頼区間の幅の関係がわかりましたか？

問題 1

(1) $n = 18 -$ (誕生月) とする．母分散は既知としたとき，標準正規分布からの n 個の標本に基づいて，信頼係数 95% と 99% の母平均の信頼区間を構成せよ．

(2) $n = 18 -$ (誕生月) とする．正規分布 $N(0, 2)$ からの n 個の標本に基づいて，信頼係数 95% と 99% の母平均の信頼区間を構成せよ．

- それぞれの問題の解説を mejiro-hanako-070112.txt に書け．
- 締め切りは 2007 年 01 月 12 日（金）13 時

ヒント 1：$\alpha\,(0 < \alpha < 1)$ を信頼係数としたとき，$k_{\alpha/2}$ を


```
qnorm(1-alpha/2,0,1)
```

と求める．

(2) のヒント

```
> k<-qt(1-alpha/2,??-1)
>
> x<-rnorm(??,0,sqrt(2))
>
> barx<-mean(x)
> sigmaU<-var(x)
> mL<-barx-k*sqrt(sigmaU/??)
> mU<-barx+k*sqrt(sigmaU/??)
```

$\widehat{\sigma}_U^2$ を

```
sigmaU<-var(x)
```

また，$t_{(n-1)}(\alpha/2)$ を


```
qt(1-alpha/2,n-1)
```

で求める．