

様々な分布を確認する

Sampo Suzuki, CC 4.0 BY-NC-SA

2021-06-29

標本標準偏差はどのような分布を取るのか？

『統計解析の話』[大平, 2006] では母集団のばらつき（母標準偏差 σ ）を推定するために母集団からサンプリングした標本の標準偏差（ s ）の分布を用いようと説明されています¹。そこで、標本標準偏差（ s ）の分布をテキストにあるように以下の手順で求めてみます。

¹ 3. 名探偵ものがたりばらつきを区間推定するには P56～

1. 標本数を $n = 2$ とする
2. 母集団²から上記の標本を取り出す
3. 標本から標本標準偏差（ s ）を計算して記録する
4. 上記を任意の回数³繰り返す
5. 記録した標本標準偏差（ s ）のヒストグラムをプロットする

² 建前上は未知の母集団

³ 本資料では 1 万回

```
1 n <- 2
2 df <- data.frame()
3 for (i in c(1:10000)) {
4   xs <- rnorm(n)
5   dftmp <- data.frame(no = i, s = sqrt(sum((xs - mean(xs))^2) / n))
6   df <- dplyr::bind_rows(df, dftmp)
7 }
```

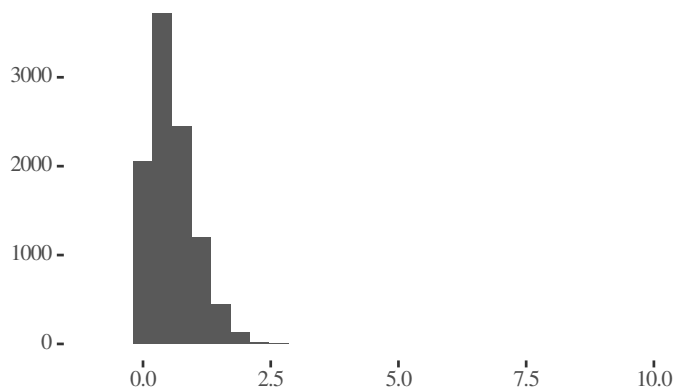
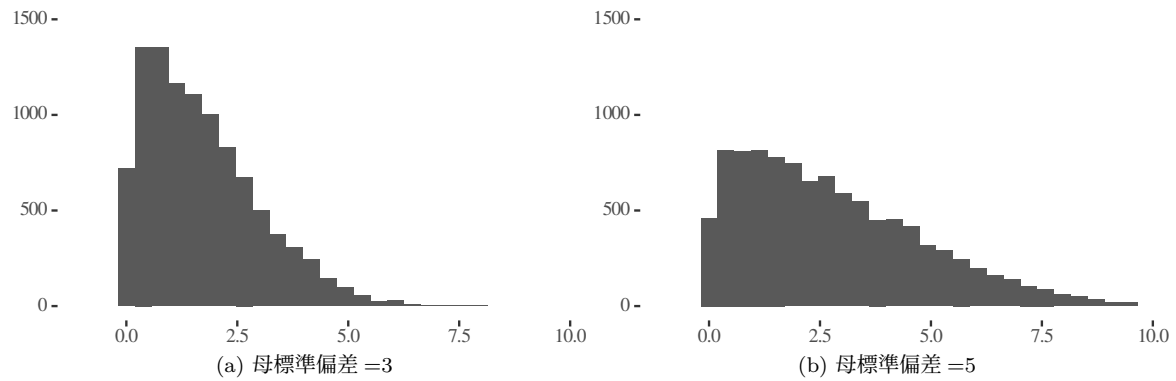


Figure 1: 標本標準偏差の分布

このように大きく右に歪んだ分布になることがわかります⁴。

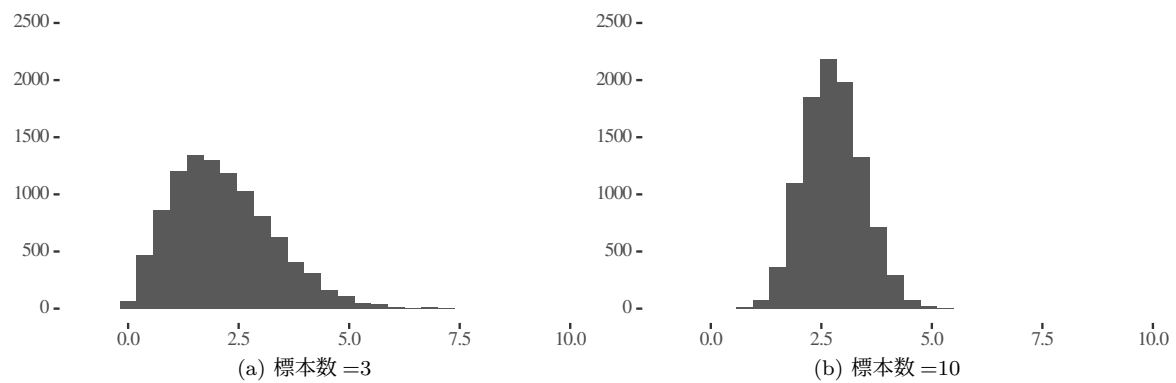
⁴ 次ページと比較しやすいように横軸をそろえてあります。縦軸は異なるので注意してください。

次に同じ手順で母集団の標準偏差 (σ) が異なると標本標準偏差 (s) の分布がどのようなようになるかを確認します。



母標準偏差 (σ) が大きくなると右側にながらかになることがわかります。

では、母標準偏差を変えずに標本数（サンプリング数）を変えた場合、どのような分布になるのでしょうか？



標本数が大きくなると母標準偏差 (σ) を中心とした正規分布に近づいていくことがわかります。

偏差平方和の分布

テキスト [大平, 2006] では計算を簡単にするために標本標準偏差ではなく偏差平方和⁵の分布を求めるようにしています。では、偏差平方和⁶の分布がどのようなになるかプロットしてみます。偏差平方和は数式を見て分かるように標本標準偏差⁷と比例関係にありますので、分布形状も比例するはずです。

⁵ 3. 名探偵ものがたりばらつきを区間推定するには P57~

⁶ $S = \sum_{i=1}^n (x_i - \bar{x})^2$

⁷ $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$

```
1 n <- 2
2 df <- data.frame()
3 for (i in c(1:10000)) {
4   xs <- rnorm(n)
5   dftmp <- data.frame(no = i, S = sum((xs - mean(xs))^2))
6   df <- dplyr::bind_rows(df, dftmp)
7 }
```

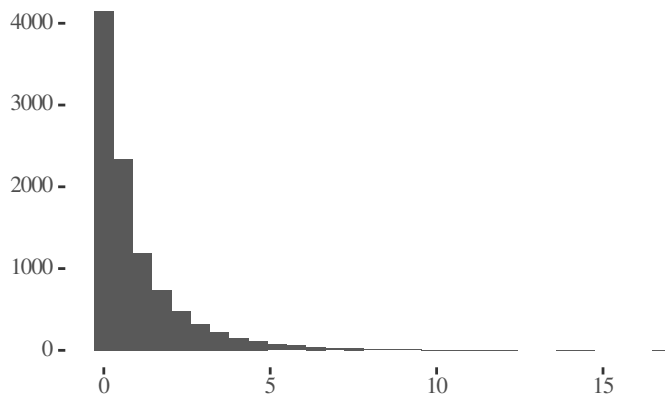
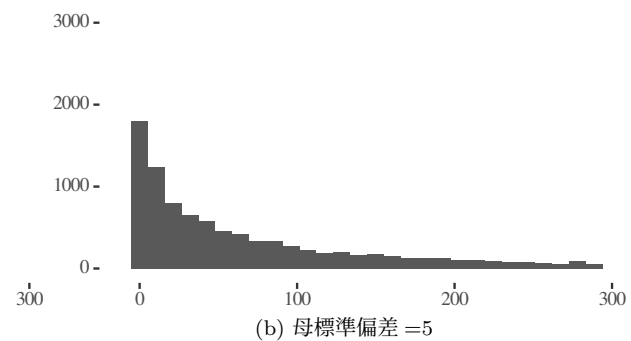
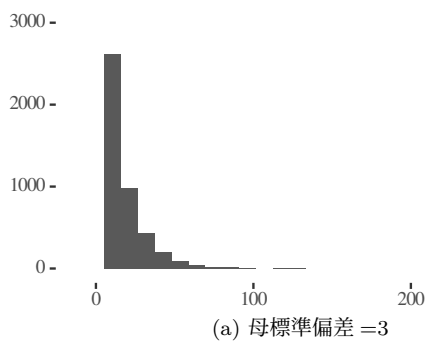


Figure 2: 偏差平方和の分布

また、テキスト⁸にあるように母標準偏差 (σ) で分布が変わるかどうかも確認します。

⁸ 3. 名探偵ものがたりばらつきを区間推定するには P58



χ^2 分布

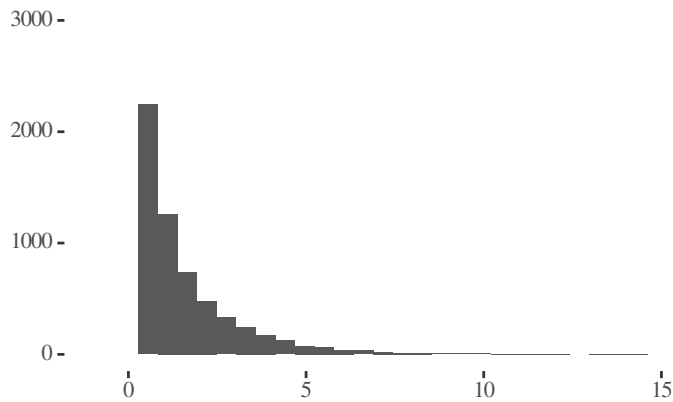
χ^2 分布⁹は標本数（自由度）により分布が変わりますが、母集団の標準偏差（ σ ）には左右されないという特徴があります。

$$^9 \chi^2 = \frac{S}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$$

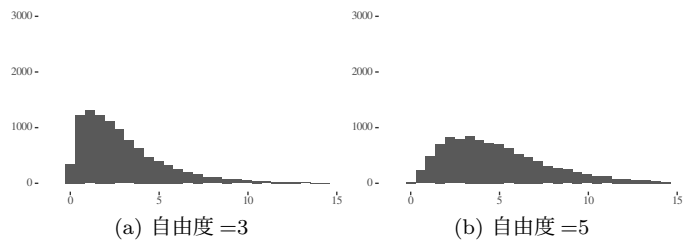
```

1  n <- 2
2  sd <- 1
3  df <- data.frame()
4  for (i in c(1:10000)) {
5    xs <- rnorm(n, sd = sd)
6    dftmp <- data.frame(no = i, S = sum((xs - mean(xs))^2), var = sd ^ 2)
7    df <- dplyr::bind_rows(df, dftmp)
8  }

```

Figure 3: 自由度=1 の χ^2 分布

自由度を変えた場合。

Figure 4: 自由度が異なる χ^2 分布

母集団の標準偏差 (σ) を変えた場合。

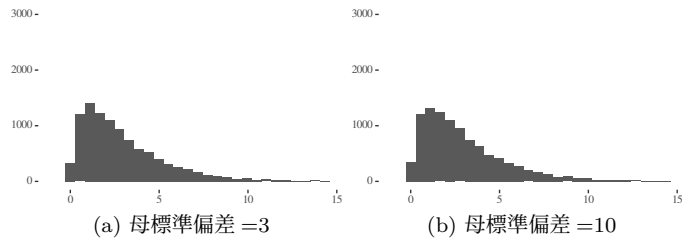


Figure 5: 自由度 = 3 の χ^2 分布

References

平大平. 『統計解析のはなし』. 日科技連出版, 改訂版 edition, 2006.
 URL <https://www.juse-p.co.jp/products/view/196>. ISBN
 978-4-8171-8028-5.