La variance empirique et la distribution khi-deux

Antonio Falcó

14/11/2019

Questions à repondre

- Sous l'hypothèse que la variable `taille` suivi une distribution normal, est-ce qu'on peut calculer l'erreur commis on prendre `var(nutriage$taille)` comme une approximation de la variance de la taille σ^2 chez la population?
- Est-ce qu'on peut trouver un intervalle $[a, b]$ tel que la probabilité de trouver la variance σ^2 à l'intervalle $[a, b]$ soit égal à $1 - \alpha$?

```r
nutriage <- read.csv("~/Descargas/Atelier4-master(2)/Atelier4-master/nutriage.csv")
names(nutriage)
```

```
##  [1] "sexe"               "situation"          "the"
##  [4] "cafe"               "taille"             "poids"
##  [7] "age"                "viande"             "poisson"
## [10] "fruit_crus"         "fruit_legume_cuits" "chocol"
## [13] "matgras"
```

```r
mu <- mean(nutriage$taille)
sigma2 <-var(nutriage$taille)
n <- length(nutriage$taille)
mu
```

```
## [1] 163.9602
```

```r
sigma2
```

```
## [1] 81.06063
```

```r
n
```

```
## [1] 226
```

Reponse

On connaît que

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

ou

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \texttt{var(X)}$$

Soit $\varepsilon > 0$ l'erreur d'approximation entre S^2et le vrai valeur chez la population σ^2:

$$|S^2 - \sigma^2| < \varepsilon$$

Alors fixée une probabilité $1 - \alpha$ pour l'erreur ε on a la suivante expression pour l'erreur d'aproximation:

$$\Pr(|S^2 - \sigma^2| < \varepsilon) = 1 - \alpha.$$

Observe que

$$
\begin{aligned}
\Pr(|S^2-\sigma^2|<\varepsilon) &= \Pr(-\varepsilon<\sigma^2-S^2<\varepsilon)\\
&= \Pr(-\varepsilon+S^2<\sigma^2<\varepsilon+-S^2)\\
&= \Pr\left(\frac{(-\varepsilon+S^2)}{(n-1)S^2}<\frac{\sigma^2}{(n-1)S^2}<(\frac{(\varepsilon+S^2)}{(n-1)S^2}\right)\\
&= \Pr\left(\frac{(n-1)S^2}{(\varepsilon+S^2)}<\frac{(n-1)S^2}{\sigma^2}<(\frac{(n-1)S^2}{(-\varepsilon+S^2)}\right)
\end{aligned}
$$

on peut concluir:

$$\Pr\left(\frac{(n-1)S^2}{(\varepsilon+S^2)}<\chi^2_{n-1}<(\frac{(n-1)S^2}{(-\varepsilon+S^2)}\right)=1-\alpha.$$

Soit

$$
\begin{aligned}
x_{\min} &= \frac{(n-1)S^2}{S^2+\epsilon}\Leftrightarrow\varepsilon=\frac{(n-1)S^2}{x_{\min}}-S^2\\
x_{\max} &= \frac{(n-1)S^2}{S^2-\epsilon}\Leftrightarrow\varepsilon=S^2-\frac{(n-1)S^2}{x_{\max}}
\end{aligned}
$$

Observe qu'il se verifié que

$$S^2-\frac{(n-1)S^2}{x_{\max}}=\frac{(n-1)S^2}{x_{\min}}-S^2\Leftrightarrow 2S^2=\frac{(n-1)S^2}{x_{\max}}+\frac{(n-1)S^2}{x_{\min}}$$

et

$$2=\frac{(n-1)}{x_{\max}}+\frac{(n-1)}{x_{\min}}\Leftrightarrow\frac{2}{n-1}=\frac{1}{x_{\min}}+\frac{1}{x_{\max}}$$

Si on prend l'expression

$$\Pr\left(x_{\min}<\chi^2_{n-1}<x_{\max}\right)=1-\alpha.$$

on peut deduir que

$$x_{\max}=\texttt{qchisq}(1-\frac{\alpha}{2},\texttt{n-1})\text{ et }x_{\min}=\texttt{qchisq}(\frac{\alpha}{2},\texttt{n-1})$$

On obtient finalement les formules suivantes:

$$\texttt{qchisq}(\frac{\alpha}{2},\texttt{n-1})=\frac{(n-1)S^2}{S^2+\epsilon}$$

et

$$\texttt{qchisq}(1-\frac{\alpha}{2},\texttt{n-1})=\frac{(n-1)S^2}{S^2-\epsilon}$$

On va resoudre le cas pratique suivant: On supposant que $\alpha=0.05$ chercher ε tel que

$$\Pr(|S^2-\sigma^2|<\varepsilon)=0.95.$$

Pour la variable `taille` on connaît que $S^2=81.06063$ et $n-1=225$. Si on calcule $x_{\min}$ et $x_{\max}$:

```
alpha = 0.05
xmin <- qchisq(alpha/2,(length(nutriage$taille)-1))
xmax <- qchisq(1-alpha/2,(length(nutriage$taille)-1))
xmin
```

```
## [1] 185.3483
```

```
xmax
```

```
## [1] 268.4378
```

On va verifier l'expresion

$$\frac{2}{n-1} = \frac{1}{x_{\min}} + \frac{1}{x_{\max}}$$

avec le logiciel:

```
(1/xmin)+(1/xmax)
```

```
## [1] 0.009120504
```

```
2/(n-1)
```

```
## [1] 0.008888889
```

En utilisant l'équation:

$$\varepsilon = S^2 - \frac{(n-1)S^2}{x_{\max}} \Leftrightarrow S^2 - \epsilon = \frac{(n-1)S^2}{x_{\max}} \Leftrightarrow \epsilon = S^2 - \frac{(n-1)S^2}{x_{\max}}$$

on peut calculer l'erreur d'aproximation à 95% de proba:

```
epsilon <- var(nutriage$taille)-(length(nutriage$taille)-1)*var(nutriage$taille)/xmax
epsilon
```

```
## [1] 13.117
```

et on obient un erreur de $\varepsilon = 13.117$.

Comme on peut concluir que

$$\Pr(|S^2 = 81.06063 - \sigma^2| < 13.117) = 0.95$$

c'est-à-dire

$$\Pr(-12.81503 < \sigma^2 - S^2 = 81.06063 < 13.117) = 0.95,$$

qui est equivalent à

$$\Pr(-13.117 + S^2 = 81.06063 < \sigma^2 < 13.117 + S^2 = 81.06063) = 0.95$$

ou $-13.117 + S^2 = 81.06063$ est egal à

```
var(nutriage$taille)-epsilon
```

```
## [1] 67.94363
```

et $13.117 + S^2 = 81.06063$ est egal à

```
var(nutriage$taille)+epsilon
```

```
## [1] 94.17763
```

Alors

$$\Pr(67.94363 < \sigma^2 < 94.17763) = 0.95$$

la variance de la `taille`, chez la population σ^2, est à l'intervalle $[67.94363, 94.17763]$ avec une probabilité de 95%.