

Distribusi dan Inferensi untuk Data Kategorik

Kurnia Susvitasari

2024-08-28

Membangkitkan Variabel Acak dari Distribusi Multinomial

Tentunya, bukan hal yang sulit bagi Anda untuk membangkitkan (bahkan mengevaluasi probabilitas/ probabilitas kumulatif/ quantil) dari variabel acak yang berdistribusi binomial. Namun, bagaimana dengan variabel acak yang berdistribusi multinomial?

Tenang saja! R menyediakan build-in function di package `stats` dengan memanggil fungsi `rmultinom`. Selain itu, Anda juga bisa mengevaluasi pmf-nya menggunakan `dmultinom`.

Proporsi Vegetarian (Parameter Binomial)

Contoh kasus ini merujuk pada buku Alan Agresti bab 1.4.3; baca [pg. 15, 1].

Dalam sebuah kuesioner yang diberikan kepada kelas pengantar statistik ($n = 25$), tidak ada siswa yang mengatakan mereka adalah vegetarian. Dengan asumsi bahwa jumlah yang merespons “ya” didistribusikan secara binomial dengan probabilitas keberhasilan π , berapakah interval kepercayaan 95% untuk π ?

Pertama, perhatikan bahwa $\hat{\pi} = 0/25 = 0$.

Dengan menggunakan Wald test, kita bisa menghitung interval kepercayaan dengan cara

```
phat <- 0
n <- 25

# 95% interval kepercayaannya adalah
phat + c(-1, 1) * qnorm(p = 0.975) * sqrt((phat * (1 - phat))/n)

## [1] 0 0
```

Alternatif lain untuk menghitung interval kepercayaan Wald adalah dengan menggunakan fungsi `Hmisc::binconf` dengan argumen `method = "asymptotic"`.

```
library(Hmisc)
binconf(x = 0, n = 25, method = "asymptotic")

## PointEst Lower Upper
##          0      0      0
```

Interval kepercayaan kedua adalah dengan menggunakan score test. Perhatikan bahwa interval kepercayaan skor mengandung nilai π_0 yang memenuhi $|S| < z_{0.025}$ dimana S diberikan di eq.(1.9). Selain itu, endpoint dari interval tersebut harus merupakan solusi dari eq.(1.10). Kita bisa menggunakan R untuk mencari solusi dari eq.(1.10) menggunakan fungsi `optim`. Tapi, untuk saat ini, mari menggunakan build-in function `prop.test` atau `Hmisc::binconf` dengan argumen `method = "wilson"`.

```
# menggunakan prop.test
score_ci <- prop.test(x = 0, n = 25, correct = F)
score_ci$conf.int
```

```
## [1] 0.0000000 0.1331923
## attr("conf.level")
## [1] 0.95
```

```
# menggunakan binconf
binconf(x = 0, n = 25, method = "wilson")
```

```
## PointEst Lower Upper
## 0 0 0.1331923
```

Selanjutnya, kita akan menghitung interval kepercayaan likelihood-ratio (LR). Sebenarnya, kita bisa saja mencari akar dari eq.(1.13) secara analitik. Namun, R memberikan kemudahan dimana kita bisa menggunakan fungsi `optim` untuk mencari akarnya.

Idenya adalah, kita ingin meminimumkan jarak kuadrat antara statistik LR dengan quantil $\chi_1^2(\alpha)$. *Jangan hanya running the code saja! Coba dimengerti!*

Berdasarkan bab 1.4.3, diperoleh $-2(L_0 - L_1) = -50 \log(1 - \pi_0) < \chi_1^2(0.05)$. Maka,

```
# fungsi objektif kita
f <- function(x, chi2) (-50*log(1-x) - chi2)^2

# interval kepercayaannya adalah
result <- optim(par = .5, fn = f, chi2 = qchisq(.95, 1), method = "L-BFGS-B", lower = 0, upper = 1)
result$par
```

```
## [1] 0.07395143
```

Menguji Teori Mendel (Parameter Multinomial)

Sebagai ilustrasi untuk menguji hipotesis pada parameter distribusi multinomial menggunakan uji khi-kuadrat Pearson dan LR, kita akan menggunakan contoh Teori Mendel; baca bab 1.5.4 [pg.19, 1].

Uji khi-kuadrat Pearson dapat dengan mudah dilakukan dengan memanggil fungsi `chisq.test`.

```
chisq.test(x = c(6022,2001), p = c(.75,.25))
```

```
##
## Chi-squared test for given probabilities
##
## data: c(6022, 2001)
## X-squared = 0.014999, df = 1, p-value = 0.9025
```

Bagaimana dengan uji khi-kuadrat LR?

```
obs <- c(6022, 2001)
expected <- 8023 * c(0.75, 0.25)
1-pchisq(2 * sum(obs * log(obs/expected)), df=1)
```

```
## [1] 0.9025024
```

Latihan

1. Coba bangkitkan 10 variabel acak berdistribusi multinomial dengan $n = 15$ dan $\pi = \{.1, .2, .8\}$.
2. Dari variabel acak yang Anda bangkitkan di no.1, cari pmf-nya untuk data ke-8.
3. Misalkan $\hat{\pi} = .2$ pada contoh Proporsi Vegetarian. Konstruksilah 95% interval kepercayaan menggunakan a) Wald, b) skor, dan c) LR.
4. Apakah interpretasi dari hasil luaran uji khi-kuadrat Pearson dan LR pada contoh Teori Mendel?
5. Mengapa hasil luaran uji khi-kuadrat Pearson dan LR sama? Dapatkah kita mengeneralisir untuk skenario lainnya?

Referensi

1. Alan Agresti. Categorical Data Analysis, Third Edition. Wiley, 2013.