02 Chi-Squared Distribution

November 20, 2022

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

1 Distribution of the empirical variance

Assume that the filling volumne X is normally distributed with expected value μ and standard deviation σ. X and σ are unknown. We estimate the unknown variance σ^2 with

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

and then the quantity

$$Q = \frac{(n-1) * s_x^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

has a known distribution. One calles it the Chi-squard (χ^2) distribution with $n-1$ degrees of freedom

2 Confidence interval for standard deviation

Let $X_1, \ldots, X_n$ be independent measurements with $X_i \sim N(\mu, \sigma^2)$, where $\theta = \sigma$ should be estimated from the data. Then

$$\hat{\theta}_L = \frac{(n-1)s_x^2}{q_{1-\frac{\alpha}{2}}}$$
$$\hat{\theta}_U = \frac{(n-1)s_x^2}{q_{\frac{\alpha}{2}}}$$

define a two-sided confidence interval with level $1-\alpha$. Here q_β is the β-quantile of the χ^2-distribution with $n-1$ dof.

[2]:
```python
std = 0.1337
x = stats.norm(1, std).rvs(100)
x[:10]
```

[2]:
```
array([0.90770125, 1.23767567, 1.04638838, 0.8994124 , 1.02956507,
       0.56225592, 1.08670374, 1.05022968, 1.1747637 , 1.00523278])
```

[3]:
```python
alpha = 0.05
n = len(x)

# Calculate the quantiles
q1 = stats.chi2(n - 1).ppf(alpha / 2)
q2 = stats.chi2(n - 1).ppf(1 - alpha / 2)

# Calculate the variance
variance = np.var(x, ddof=1)

# Calculate confidence interval
lo = (n - 1) * variance / q2
up = (n - 1) * variance / q1

print(f"Confidence interval: [{lo}, {up}]")
```

```
Confidence interval: [0.013944984549628035, 0.024411345127434987]
```

[4]:
```python
print(f"Correct variance: {std ** 2}")
```

```
Correct variance: 0.017875690000000003
```