A33 Maximum-Likelihood

Samstag, 13. November 2021 12:02

David Venker
Jan Jäkel
Nico Guth

Aufgabe 33 *Maximum-Likelihood*

Eine Zufallsvariable x soll einer Gleichverteilung

$$f(x) = \begin{cases} 1/b & 0 \leq x \leq b \\ 0 & x < 0 \quad \text{oder} \quad x > b \end{cases}$$

folgen.

(a) Bestimmen Sie einen Schätzer für den Parameter b mit der Maximum Likelihood Methode aus einer Stichprobe $x_1, x_2, \dots, x_n$.

Likelihood Funktion:

$$L = \prod_{i=1}^{n} f(x_i) \qquad f(x_i) = \frac{1}{b} \, \mathbb{1}_{[0,b]}(x_i)$$

$$\text{wobei} \quad \mathbb{1}_{[0,b]}(x) = \begin{cases} 1 & x \in [0,b] \\ 0 & x \notin [0,b] \end{cases}$$

$$L(\vec{x}) = \frac{1}{b^n} \underbrace{\prod_{i=1}^{n} \mathbb{1}_{[0,b]}(x_i)}_{\mathbb{1}_{[0,b]}(\vec{x}) :=}$$

$$L(b|\vec{x}) = \frac{1}{b^n} \mathbb{1}_{[0,b]}(\vec{x}) \qquad \text{wobei} \quad \mathbb{1}_{[0,b]}(\vec{x}) = \begin{cases} 1 & \text{wenn alle } x_i \in [0,b] \\ 0 & \text{sonst} \end{cases}$$

- maximiere $L(b|\vec{x})$:

1) $b \overset{!}{\geq} \max\{x_i\}$ sonst $L(b|\vec{x}) = 0$, wobei $\{x_i\}$ die Menge aller Messwerte ist

2) $\frac{1}{b^n}$ ist streng monoton fallend

Also wird $L(b|\vec{x})$ maximal für:

$$\hat{b} = \max\{x_i\}$$

(b) Ist diese Schätzung erwartungstreu? Wenn nein, wie kann das in diesem Fall korrigiert werden?

Erwartungstreu: $E(\hat{b}) \overset{!}{=} b$

$$\Leftrightarrow \quad E(\hat{b}) = E(\max\{x_i\}) = \int_{-\infty}^{\infty} y \, \mathrm{PDF}(\max\{x_i\} = y) \, dy$$

$\Leftrightarrow E(\hat{b}) = E(\max\{x_i\}) = \int_{-\infty} y \, PDF(\max\{x_i\} = y) \, dy$

Dafür gesucht ist $P(\max\{x_i\} = y)$. (als W'keitsdichte)

Also die W'keit, dass der maximale Wert einer Stichprobe aus n Werten y ist.
Wobei die x_i gleichverteilt im Bereich $[a,b]$ sind.

Herleitung über die CDF einer Gleichverteilung zwischen $[a,b]$:

W'keit, dass ein Messwert $x_i \leq x$:

$$P(X_i \leq x) = CDF_{uniform}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b < x \end{cases}$$

W'keit, dass alle $x_i \leq y$ sind, ist gleichbedeutend mit Wkeit für $\max\{x_i\} \leq y$.
also CDF der gesuchten PDF:

$$P(\max\{x_i\} \leq y) = \begin{cases} 0 & y < a \\ (P(X_i \leq y))^n & a \leq y \leq b \\ 1 & b < y \end{cases}$$

Wir erhalten die $PDF(y) = \frac{d}{dy} P(\max\{x_i\} \leq y)$

$$PDF(\max\{x_i\} = y) = \begin{cases} 0 & y < a, \ y > b \\ n \frac{y^{n-1}}{(b-a)^n} & a \leq y \leq b \end{cases}$$

$$\Rightarrow E(\hat{b}) = \int_{-\infty}^{\infty} y \, PDF(\max\{x_i\} = y) \, dy$$

$$= \int_0^b y \, n \frac{y^{n-1}}{b^n} \, dy = \left[\frac{n}{n+1} \cdot \frac{1}{b^n} \cdot y^{n+1}\right]_0^b$$

$$= \frac{n}{n+1} \cdot \frac{1}{b^n} b^{n+1}$$

$$= \frac{n}{n+1} b$$

$\Rightarrow$ Der Schätzer $\hat{b} = \max\{x_i\}$ ist nicht erwartungstreu.

Korrektur finden:

Korrektur finden:

$$E(\hat{b}) = \frac{n}{n+1} b$$

$$\Longleftrightarrow \frac{n+1}{n} E(\hat{b}) = b$$

$$\Longleftrightarrow E\left(\frac{n+1}{n} \hat{b}\right) = b$$

$\Rightarrow$ korrigierter Schätzer: $\hat{b}' = \frac{n+1}{n} \max\{x_i\}$

Aufgabe 34 *Likelihoodkurve*

Aus einer Poisson-Verteilung werden 3 Stichproben, nämlich die Zahlen 13, 8 und 9 entnommen.

(a) Berechnen Sie die negative Log-Likelihood-Funktion als Funktion des einzigen Parameters λ und stellen Sie sie graphisch dar.

Poisson-Verteilung: $P(k|\lambda) = \frac{\lambda^k}{k!} \exp(-\lambda)$ $\lambda = n \cdot p$

$n \hat{=}$ Anzahl der Versuche
$p \hat{=}$ Erfolgsw'keit
$k \hat{=}$ Anzahl der Erfolge

gegeben: $N = 3$ unabhängige Messwerte k_i
$(k_1 = 13, k_2 = 8, k_3 = 9)$

Likelihood-Funktion:

$$L(\lambda|\vec{k}) = \prod_{i=1}^{N} P(k_i|\lambda)$$

$$F(\lambda|\vec{k}) := -\ln(L(\lambda|\vec{k}))$$

hier:

$$\Rightarrow F(\lambda|\vec{k}) = -\ln(P(k_1|\lambda)) - \ln(P(k_2|\lambda)) - \ln(P(k_3|\lambda))$$

mit $\ln(P(k_i|\lambda)) = \ln\left(\frac{\lambda^{k_i}}{k_i!} \exp(-\lambda)\right) = k_i \ln(\lambda) - \ln(k_i!) - \lambda$

$$\Leftrightarrow F(\lambda|\vec{k}) = -\sum_{i=1}^{N} (k_i \ln(\lambda) - \ln(k_i!) - \lambda)$$

$$\Leftrightarrow \boxed{F(\lambda|\vec{k}) = +N\lambda - \ln(\lambda) \sum_{i=1}^{N} k_i + \sum_{i=1}^{N} \ln(k_i!)}$$

mit $k_1 = 13$, $k_2 = 8$, $k_3 = 9$:

$$F(\lambda|\vec{k}) = 3\lambda - \ln(\lambda) \cdot 30 + 45{,}9586$$

(b) Bei welchem Wert von λ liegt das Minimum $-\ln \mathcal{L}_{\text{max}}$?

Minimum bestimmen mit $\frac{d}{d\lambda} F(\lambda|\vec{k}) \overset{!}{=} 0$

Minimum bestimmen mit $\frac{d}{d\lambda} F(\lambda|\vec{k}) \overset{!}{=} 0$

$$\frac{d}{d\lambda} F(\lambda|\vec{k}) = N - \frac{1}{\lambda}\sum_{i=1}^{N} k_i = 3 - \frac{30}{\lambda} \overset{!}{=} 0$$

$\Leftrightarrow \hat{\lambda} = \frac{1}{N}\sum_{i=1}^{N} k_i$ ist der Schätzer für λ

$\Leftrightarrow \hat{\lambda} = 10$

(c) Für welche Werte von λ nimmt $-\ln \mathcal{L}$ die Werte

$$-\ln \mathcal{L}_{\text{max}} + \frac{1}{2} \quad ,$$
$$-\ln \mathcal{L}_{\text{max}} + 2 \quad \text{und}$$
$$-\ln \mathcal{L}_{\text{max}} + \frac{9}{2}$$

an und was sagen diese Werte aus?

$F(\lambda|\vec{k}) \overset{!}{=} F(\hat{\lambda}|\vec{k}) + c$ wobei $c \in \{\frac{1}{2}, 2, \frac{9}{2}\}$

$$F(\hat{\lambda}|\vec{k}) = 3\cdot 10 - \ln(10)\cdot 30 + 45{,}9586$$
$$\approx 6{,}8810$$

$\Rightarrow 3\lambda - \ln(\lambda)\cdot 30 + 45{,}9586 \overset{!}{=} F(\hat{\lambda}|\vec{k}) + c$

nicht analytisch lösbar

(d) Vergleichen Sie diese Werte mit der Näherung über eine Taylor-Entwicklung 2. Ordnung, indem Sie die Näherung sowohl zusammen mit der Likelihood graphisch darstellen als auch die Werte aus **(c)** bestimmen. Wofür könnte die Näherung nützlich sein?

$$F(\lambda|\vec{k}) = +N\lambda - \ln(\lambda)\sum_{i=1}^{N} k_i + \sum_{i=1}^{N} \ln(k_i!)$$

$$F(\lambda|\vec{k}) = +N\lambda - \ln(\lambda)\sum_{i=1}^{N} k_i + \sum_{i=1}^{N}\ln(k_i!)$$

Taylor-Entwicklung von $F(\lambda|\vec{k})$ um $\hat{\lambda}$ bis zur 2. Ordnung:

$$F(\lambda|\vec{k}) \approx F(\hat{\lambda}) + F'(\hat{\lambda})(\lambda-\hat{\lambda}) + \frac{1}{2}F''(\hat{\lambda})(\lambda-\hat{\lambda})^2$$

$$F(\lambda|\vec{k}) \approx F(\hat{\lambda}) + \left(N - \frac{1}{\hat{\lambda}}\sum_{i=1}^{N} k_i\right)(\lambda-\hat{\lambda}) + \frac{1}{2}\left(+\frac{1}{\hat{\lambda}^2}\sum_{i=1}^{N} k_i\right)(\lambda-\hat{\lambda})^2$$

Damit die analytische Lösung von c):

$$F(\lambda) \overset{!}{=} F(\hat{\lambda}) + c$$

$$\Leftrightarrow \left(N - \frac{1}{\hat{\lambda}}\sum_{i=1}^{N} k_i\right)(\lambda-\hat{\lambda}) + \frac{1}{2}\left(+\frac{1}{\hat{\lambda}^2}\sum_{i=1}^{N} k_i\right)(\lambda-\hat{\lambda})^2 - c = 0$$

definiere $\alpha := \sum_{i=1}^{N} k_i$

$$\Leftrightarrow \left(N - \frac{\alpha}{\hat{\lambda}}\right)(\lambda-\hat{\lambda}) + \frac{1}{2}\frac{\alpha}{\hat{\lambda}^2}(\lambda-\hat{\lambda})^2 - c = 0$$

$$\Leftrightarrow \frac{1}{2}\frac{\alpha}{\hat{\lambda}^2}\lambda^2 + \lambda\left(N - \frac{\alpha}{\hat{\lambda}} - \frac{\alpha}{\hat{\lambda}^2}\hat{\lambda}\right) - \hat{\lambda}\left(N - \frac{\alpha}{\hat{\lambda}}\right) + \frac{1}{2}\frac{\alpha}{\hat{\lambda}^2}\hat{\lambda}^2 - c = 0$$

$$\Leftrightarrow \lambda^2 + \frac{2\hat{\lambda}^2}{\alpha}\lambda\left(N - 2\frac{\alpha}{\hat{\lambda}}\right) - 2\frac{\hat{\lambda}^2}{\alpha}\hat{\lambda}\left(N - \frac{\alpha}{\hat{\lambda}}\right) + 2\frac{\hat{\lambda}^2}{\alpha}\frac{1}{2}\frac{\alpha}{\hat{\lambda}^2}\hat{\lambda}^2 - 2\frac{\hat{\lambda}^2}{\alpha}c = 0$$

$$\Leftrightarrow \lambda^2 + \lambda\left(2\frac{\hat{\lambda}^2}{\alpha}N - 4\hat{\lambda}\right) - \frac{2}{\alpha}\hat{\lambda}^3\left(N - \frac{\alpha}{\hat{\lambda}}\right) + \hat{\lambda}^2 - 2\frac{\hat{\lambda}^2}{\alpha}c = 0$$

pq-Formel

$$\Rightarrow \lambda_{1,2} = -\left(\frac{\hat{\lambda}^2}{\alpha}N - 2\hat{\lambda}\right) \pm \left[\left(\frac{\hat{\lambda}^2}{\alpha}N - 2\hat{\lambda}\right)^2 + \frac{2}{\alpha}\hat{\lambda}^3\left(N - \frac{\alpha}{\hat{\lambda}}\right) - \hat{\lambda}^2 + 2\frac{\hat{\lambda}^2}{\alpha}c\right]^{\frac{1}{2}}$$

blatt15

November 14, 2021

1 Aufgabe 34 Likelihoodkurve (numerischer Teil)

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import math
```

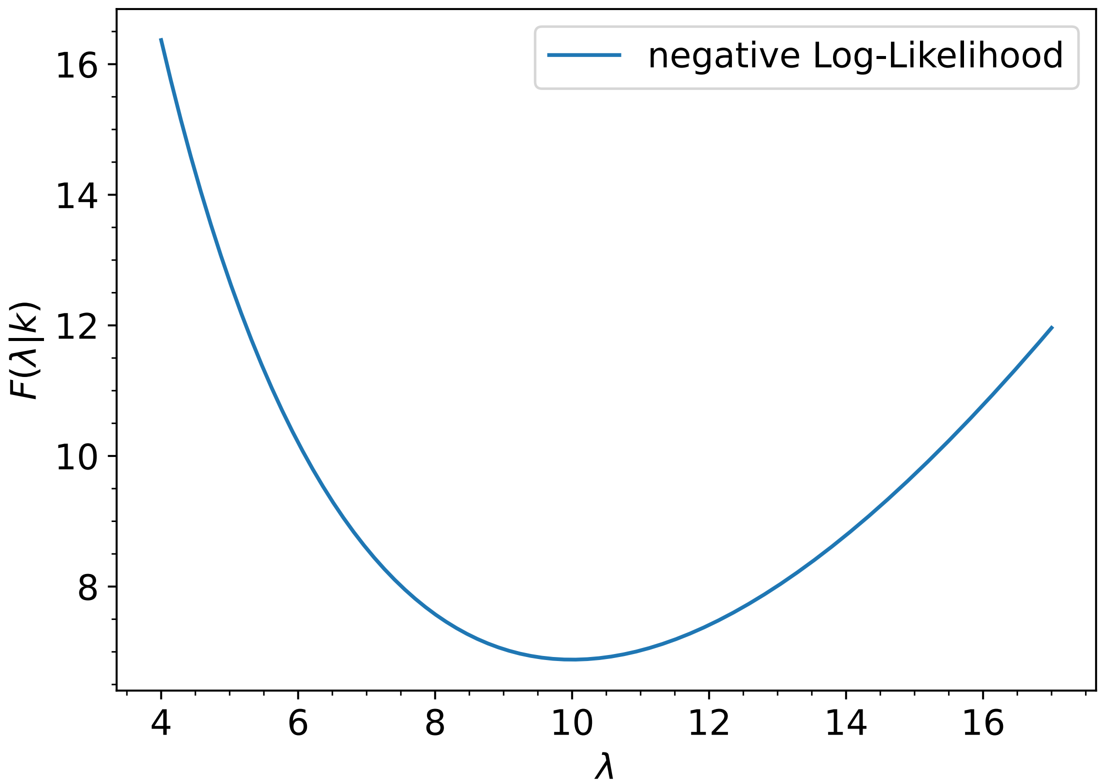
[2]:
```
# matplotlib Einstellungen
%config InlineBackend.figure_formats = ['svg']
import matplotlib as mpl
mpl.rcParams['font.size'] = 14
mpl.rcParams['figure.figsize'] = (7,5)
mpl.rcParams['xtick.minor.visible'] = True
mpl.rcParams['ytick.minor.visible'] = True
```

1.1 a) Plot der negativen Log-Likelihood-Funktion

[3]:
```
k = [13,8,9]
```

[4]:
```
# negative Log-Likelihood
def F(lam, k):
    return len(k)*lam - np.log(lam)*np.sum(k) + np.sum(np.log([math.
 ↪factorial(k_i) for k_i in k]))
```

[5]:
```
# plotten
lam = np.linspace(4,17, 10000)
plt.plot(lam, F(lam, k), label='negative Log-Likelihood')
plt.xlabel(r'$\lambda$')
plt.ylabel(r'$F(\lambda | k)$')
plt.legend()
plt.show()
```



1.2 c) Schnittstellen mit $c \in \{\frac{1}{2}, 2, \frac{9}{2}\}$ bestimmen

[6]:
```
from scipy.optimize import root
```

[7]:
```
lam0 = 10 # aus b)
c = [1/2, 2, 9/2]
```

[8]:
```
# Schnittstellen bestimmen mit F(lam) - F(lam0) - c = 0
lam_left = []
lam_right = []
for c_i in c:
    res = root(lambda x: (F(x,k) - F(lam0,k) - c_i) , x0=[lam0-3,lam0+3])
    lam_left.append(res.x[0])
    lam_right.append(res.x[1])

print('Werte für Lambda:')
df = pd.DataFrame({'lower': lam_left, 'higher': lam_right}, index=[f'c={c_i}'␣
↪for c_i in c])
df
```

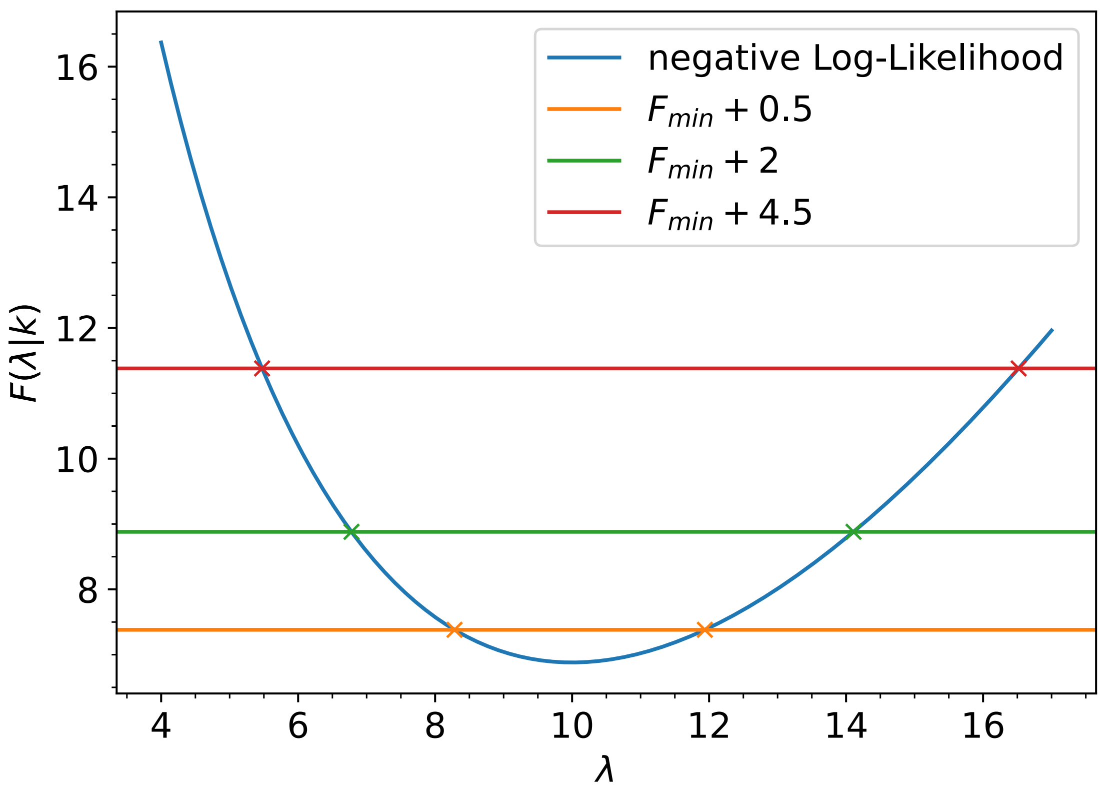
```
Werte für Lambda:
```

[8]:
```
           lower     higher
c=0.5  8.283637  11.938503
c=2    6.778765  14.108810
c=4.5  5.473703  16.519662
```

[9]:
```python
# plotten
lam = np.linspace(4,17, 10000)
plt.plot(lam, F(lam, k), label='negative Log-Likelihood')

colors = plt.rcParams["axes.prop_cycle"].by_key()["color"]
colors.pop(0)
for c_i,color,left,right in zip(c,colors,lam_left,lam_right):
    plt.axhline(F(lam0,k)+c_i, color=color, label=r'$F_{min}'+f' + {c_i}$')
    plt.plot(left, F(lam0,k)+c_i, 'x', color=color)
    plt.plot(right, F(lam0,k)+c_i, 'x', color=color)

plt.xlabel(r'$\lambda$')
plt.ylabel(r'$F(\lambda | k)$')
plt.legend()
plt.show()
```



1.2.1 Was sagen diese λ Werte aus?

Die Schnittstellen markieren die 1σ, 2σ und 3σ Umgebungen unserer Schätzung für λ.

1.3 d) Taylor-Entwicklung plotten und Schnittstellen aus c) berechnen

[10]:
```python
# Taylor-Entwicklung von F(lam)
def F_Taylor(lam,k, lam0):
    a = np.sum(k) # alpha
    N = len(k)
    return F(lam0,k) + (N-a/lam0)*(lam-lam0) + 1/2*a/lam0**2*(lam-lam0)**2
```

[11]:
```python
# analytische Nullstellen von F(lam) - F(lam0) - c = 0 , mit der Taylor-Entwicklung
def lam1(k,lam0,c):
    a = np.sum(k) # alpha
    N = len(k)
    return -(lam0**2*N/a - 2*lam0) + np.sqrt( (lam0**2*N/a - 2*lam0)**2 +2*lam0**3/a*(N-a/lam0) - lam0**2 + 2*lam0**2/a*c )

def lam2(k,lam0,c):
    a = np.sum(k) # alpha
    N = len(k)
    return -(lam0**2*N/a - 2*lam0) - np.sqrt( (lam0**2*N/a - 2*lam0)**2 +2*lam0**3/a*(N-a/lam0) - lam0**2 + 2*lam0**2/a*c )
```

[12]:
```python
# genäherte Schnittstellen berechnen
lam_left2 = []
lam_right2 = []
for c_i in c:
    lam_left2.append(lam2(k,lam0,c_i))
    lam_right2.append(lam1(k,lam0,c_i))

print('Schnittpunkte (Lambda):')
df['lower (Taylor)'] = lam_left2
df['higher (Taylor)'] = lam_right2
df
```

```
Schnittpunkte (Lambda):
```

[12]:

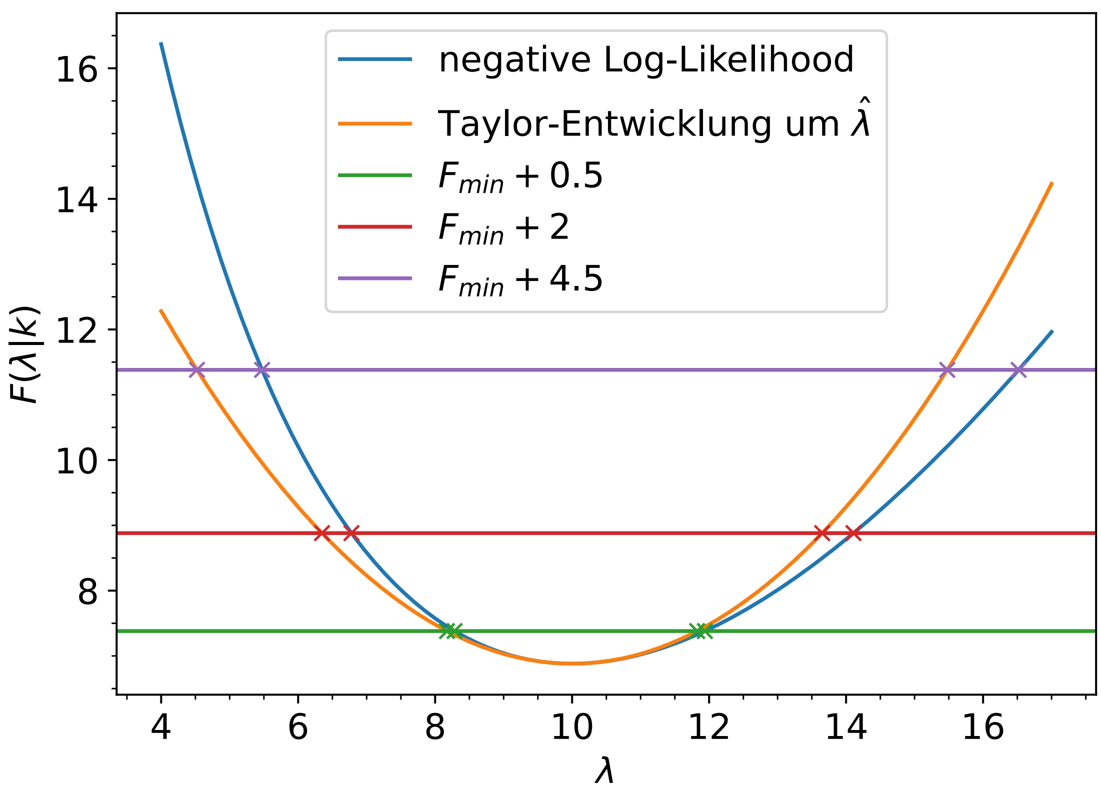
	lower	higher	lower (Taylor)	higher (Taylor)
c=0.5	8.283637	11.938503	8.174258	11.825742
c=2	6.778765	14.108810	6.348516	13.651484
c=4.5	5.473703	16.519662	4.522774	15.477226

[13]:
```python
# plotten
lam = np.linspace(4,17, 10000)
plt.plot(lam, F(lam, k), label='negative Log-Likelihood')
```

```
plt.plot(lam, F_Taylor(lam, k, lam0), label=r'Taylor-Entwicklung um␣
↪$\hat{\lambda}$')

colors = plt.rcParams["axes.prop_cycle"].by_key()["color"]
colors.pop(0)
colors.pop(0)
for c_i,color,left,right,left2,right2 in␣
↪zip(c,colors,lam_left,lam_right,lam_left2,lam_right2):
    plt.axhline(F(lam0,k)+c_i, color=color, label=r'$F_{min}'+f' + {c_i}$')
    plt.plot(left, F(lam0,k)+c_i, 'x', color=color)
    plt.plot(right, F(lam0,k)+c_i, 'x', color=color)
    plt.plot(left2, F(lam0,k)+c_i, 'x', color=color)
    plt.plot(right2, F(lam0,k)+c_i, 'x', color=color)

plt.xlabel(r'$\lambda$')
plt.ylabel(r'$F(\lambda | k)$')
plt.legend()
plt.show()
```



1.3.1 Vergleich der Näherung mit der ungenäherten negativen Log-Likelihood-Funktion:

Wie bei Taylor-Entwicklungen üblich, wird die Näherung immer ungenauer, je weiter man sich vom Minimum entfernt.
Allerdings werden kleinere Werte unterschätzt und größere Werte überschätzt.
Die exakte Funktion ist asymmetrisch um das Minimum und die Taylor-Entwicklung 2ter Ordnung ist eine (symmetrische) Parabel.

1.3.2 Wofür könnte die Näherung nützlich sein?

Die 1σ Umgebung der Schätzung wird immernoch recht genau berechnet.
Eine solche Näherung könnte den Rechenaufwand deutlich verringern, da ansonsten die Schnittstellen numerisch bestimmt werden müssen.

```
[ ]:
```