David Venker
Jan Jäkel
Nico Guth

Aufgabe 37 *γ-Astronomie 2*

Bei dieser Aufgabe handelt es sich um eine Fortführung der Aufgabe *γ-Astronomie* vom letzten Blatt. Jetzt soll festgestellt werden ob sich an der Position, auf die das Teleskop gerichtet war, wirklich eine γ-Quelle befindet. Hierzu wird die Nullhypothese verwendet. Zur Erinnerung die Likelihoodfunktion lautete

$$\ln L = -F = N_{\text{off}} \ln(b) + N_{\text{on}} \ln(s + \alpha b) - (1+\alpha)b - s - \ln(N_{\text{off}}!) - \ln(N_{\text{on}}!) \tag{1}$$

und folgende Werte für s und b machten diese Likelihood maximal:

$$\hat{s} = N_{\text{on}} - \alpha N_{\text{off}} \tag{2}$$
$$\hat{b} = N_{\text{off}} \tag{3}$$

(a) Die Nullhypothese besagt, dass es gar keine γ-Quelle gibt, also $s_0 = 0$. Welcher Wert und welcher Fehler ergeben sich unter dieser Annahme für b_0 nach der Methode der maximalen Likelihood?

Likelihood: $\alpha = \frac{t_{on}}{t_{off}}$

$$F(s,b \mid N_{on}, N_{off}) = -N_{off} \ln(b) - N_{on} \ln(s+\alpha b) + (1+\alpha) b + s + \ln(N_{off}!) + \ln(N_{on}!)$$

Nullhypothese: $s_0 = 0$

Welches b_0 ergibt sich dafür?

gesucht ist das Maximum von $-F$ wenn $s_0 = 0$

$$\frac{dF}{db} = -\frac{N_{off}}{b} - \frac{N_{on}}{b} + (1+\alpha) \stackrel{!}{=} 0$$

$$\Rightarrow \hat{b}_0 = \frac{N_{off} + N_{on}}{1+\alpha}$$

Varianz:

$$\frac{d^2 F}{db^2} = + \frac{N_{off}}{b^2} + \frac{N_{on}}{b^2} = + \frac{1}{b^2}(N_{off} + N_{on})$$

$$Var(\hat{b}_0) = \left(\frac{1}{\frac{d^2F}{db^2}\Big|_{\hat{b}_0}} \right) = + \hat{b}_0^2 (N_{off} + N_{on})^{-1}$$

$$= \left(\frac{N_{off} + N_{off}}{1+\alpha} \right)^2 (N_{off} + N_{on})^{-1}$$

$$Var(\hat{b}_0) = \frac{N_{off} + N_{on}}{(1+\alpha)^2}$$

$$\sigma = \sqrt{Var(\hat{b}_0)} = \frac{\sqrt{N_{off} + N_{on}}}{1+\alpha}$$

(b) Wie lautet das Verhältnis λ der beiden Likelihoods?

Likelihood:

$$L(s,b \mid N_{on}, N_{off}) = \exp(-F(s,b \mid N_{on}, N_{off}))$$

Likelihood:

$$L(s,b\,|\,N_{on},N_{off}) = \exp(-F(s,b\,|\,N_{on},N_{off}))$$

$$= \exp\left(+N_{off}\ln(b) + N_{on}\ln(s+\alpha b) - (1+\alpha)b - s - \ln(N_{off}!) - \ln(N_{on}!)\right)$$

$$= b^{N_{off}} \cdot (s+\alpha b)^{N_{on}} \cdot \exp(-(1+\alpha)b - s) \cdot \frac{1}{N_{off}!} \cdot \frac{1}{N_{on}!}$$

$$\lambda = \frac{L(s_0,\hat{b}_0\,|\,N_{off},N_{on})}{L(\hat{s},\hat{b}\,|\,N_{off},N_{on})} \quad \text{mit} \quad \hat{s} = N_{on} - \alpha N_{off} \qquad s_0 = 0$$

$$\hat{b} = N_{off} \qquad \hat{b}_0 = \frac{N_{off}+N_{on}}{1+\alpha}$$

$$\lambda = \frac{\hat{b}_0^{N_{off}} \cdot (s_0+\alpha\hat{b}_0)^{N_{on}} \cdot \exp(-(1+\alpha)\hat{b}_0 - s_0) \cdot \cancel{\frac{1}{N_{off}!}} \cdot \cancel{\frac{1}{N_{on}!}}}{\hat{b}^{N_{off}} \cdot (\hat{s}+\alpha\hat{b})^{N_{on}} \cdot \exp(-(1+\alpha)\hat{b} - \hat{s}) \cdot \cancel{\frac{1}{N_{off}!}} \cdot \cancel{\frac{1}{N_{on}!}}}$$

$$= \frac{\left(\frac{N_{off}+N_{on}}{1+\alpha}\right)^{N_{off}} \cdot \left(\alpha\frac{N_{off}+N_{on}}{1+\alpha}\right)^{N_{on}} \exp\left(-(1+\alpha)\left(\frac{N_{off}+N_{on}}{1+\alpha}\right)\right)}{N_{off}^{N_{off}} \cdot \left((N_{on}-\alpha N_{off})+\alpha N_{off}\right)^{N_{on}} \exp\left(-(1+\alpha)N_{off} - (N_{on}-\alpha N_{off})\right)}$$

$$= \frac{\left(\frac{N_{off}+N_{on}}{1+\alpha}\right)^{N_{off}} \cdot \left(\alpha\frac{N_{off}+N_{on}}{1+\alpha}\right)^{N_{on}} \exp(-(N_{off}+N_{on}))}{N_{off}^{N_{off}} \cdot N_{on}^{N_{on}} \exp(-(N_{off}+N_{on}))}$$

$$\lambda = \frac{(N_{off}+N_{on})^{N_{off}+N_{on}}}{N_{off}^{N_{off}} \cdot N_{on}^{N_{on}}} \cdot \frac{\alpha^{N_{on}}}{(1+\alpha)^{N_{off}+N_{on}}}$$

(c) Unter den gegebenen Hypothesen und mit großen N_{on}, N_{off} ist $D = -2\ln\lambda$ χ^2-verteilt mit einem Freiheitsgrad. Mit welcher Konfidenz lehnen Sie die Nullhypothese ab? Geben Sie Ihr Ergebnis in Einheiten von Sigma an.

Tipp: Betrachten Sie eine standardnormalverteilte Variable u. Welcher Verteilung folgt u^2? Vergleichen Sie mit D.

$D = -2\ln(\lambda)$ ist χ^2 verteilt mit einem Freiheitsgrad

$$= -2\ln\left(\frac{(N_{off}+N_{on})^{N_{off}+N_{on}}}{N_{off}^{N_{off}} \cdot N_{on}^{N_{on}}} \cdot \frac{\alpha^{N_{on}}}{(1+\alpha)^{N_{off}+N_{on}}}\right)$$

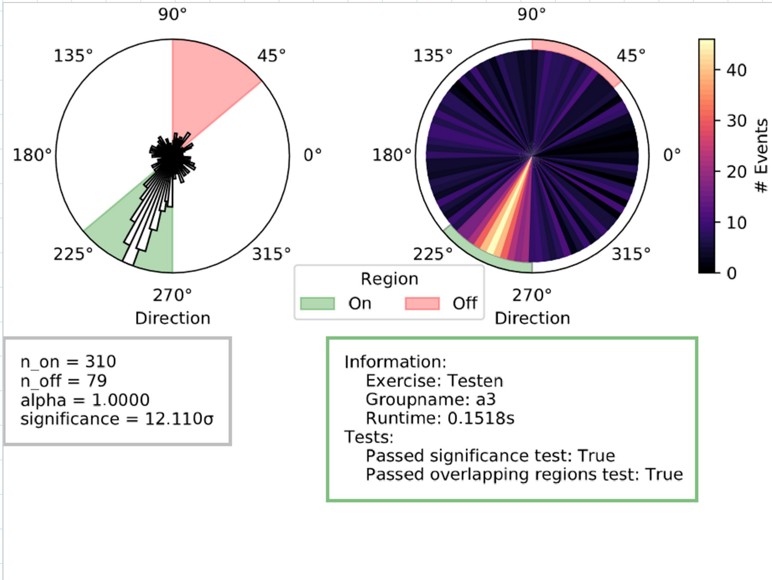
$$= -2\left[(N_{off}+N_{on})\ln(N_{off}+N_{on}) - N_{off}\ln(N_{off}) - N_{on}\ln(N_{on}) + N_{on}\ln(\alpha) - (N_{off}+N_{on})\ln(1+\alpha)\right]$$

... (keine Ahnung)

Signifikanz: $\alpha = \sqrt{D} \cdot \sigma$

(d) Für diesen Aufgabenteil sollen Sie wieder das Projekt-Repository aus dem letzten Semester verwenden. Nutzen Sie dieses, um einen Datensatz mit 10000 Ereignissen zu erstellen. Der Datensatz soll auch für spätere Aufgaben verwendet werden.

N_off Region: 220° bis 270°
N_on Region: 40° bis 90°



- Was passiert, wenn Sie die Untergrundregion stark vergrößern?
    - Die Signifikanz wird größer. Aber auch alpha wird größer.
- Was passiert bei einer höheren Anzahl von Ereignissen?
    - Die Signifikanz wird größer. Der Peak wird schmaler.

- Welches Problem sehen Sie, wenn man in der laufenden Analyse die Quellregion beliebig ändert?
    - Man sollte die Nullhypothese vor der Analyse aufstellen.
    - Wenn man die Region so wählt, dass die Signifikanz am größten wird, dann führt man einen Bias ein und die Analyse ist nicht mehr unvoreingenommen.

stats.py ×

project_a3 > stats.py > ...

```python
import numpy as np


def significance(n_on, n_off, alpha):
    """Calculate the significance of an observation.

    Parameters
    ----------
    n_on : int
        Count of events in On-Region.
    n_off : int
        Count of events in Off-Region.
    alpha : float
        Weighting of Off-Region.

    Returns
    -------
    float
    """

    D = -2* ( (n_off + n_on)*np.log(n_off+n_on) - n_off*np.log(n_off) - n_on*np.log(n_on) + n_on*np.log(alpha) - (n_off+n_on)*np.log(1+alpha) )

    return np.sqrt(D)


def test_significance():
```

```
Run Cell | Run Above | Debug Cell
# %%
import numpy as np
Run Cell | Run Above | Debug Cell
# %%
!python exercises/testen.py a3 {data_dir} -d {data_dir} \
    --on-region {np.deg2rad(220)},{np.deg2rad(270)} \
    --off-region {np.deg2rad(40)},{np.deg2rad(90)}
```

Aufgabe 38 *Zwei Histogramme*

Gegeben sind zwei Histogramme mit dem gleichen Binning (r Bins). Die Hypothese ist, dass beide Histogramme Zufallszahlen mit der gleichen Verteilung repräsentieren. Das heißt es existieren r Wahrscheinlichkeiten $p_1, \dots, p_r$ dafür, dass eine Beobachtung im i-ten Bin liegt ($\sum_{i=1}^r p_i = 1$). Die Einträge im i-ten Bin des ersten Histogramm werden als n_i und im zweiten m_i bezeichnet. Die Anzahl der Beobachtungen im ersten Histogramm ist $N = \sum_{i=1}^r n_i$ und im zweiten $M = \sum_{i=1}^r m_i$.

(a) Welcher Verteilung folgen die Zählraten in den einzelnen Bins? Stellen sie die PDF für ein einzelnes Bin für beide Histogramme (n_i und m_i) unter Annahme der Nullhypothese auf.

Es handelt sich um Zählraten → Poisson-Verteilung : $PDF(k) = \frac{\lambda^k}{k!} \exp(-\lambda)$

$$\langle n_i \rangle = p_i \cdot N \quad \Rightarrow \quad PDF(n_i) = \frac{1}{n_i!} \cdot \langle n_i \rangle^{n_i} \exp(-\langle n_i \rangle)$$

$$= \frac{1}{n_i!} (p_i N)^{n_i} \exp(-p_i N) \quad \text{mit} \quad \sum_{i=1}^{r} n_i \overset{!}{=} N$$

analog

$$\langle m_i \rangle = p_i \cdot M \quad \Rightarrow \quad PDF(m_i) = \frac{1}{m_i!} (p_i \cdot M)^{m_i} \exp(-p_i M) \quad \text{mit} \quad \sum_{i=1}^{r} m_i \overset{!}{=} M$$

(b) Stellen sie die Likelihood Funktion für die Nullhypothese auf. Finden sie den Schätzer $\hat{p}_i$, der die Likelihood maximiert.

$$L(p_i | n_i, m_i) = P(n_i | p_i) \cdot P(m_i | p_i) = PDF(n_i) \cdot PDF(m_i)$$

negative Log-Likelihood: $F(p_i) := -\ln(L(p_i | n_i, m_i)) = -\ln(PDF(n_i)) - \ln(PDF(m_i))$

$$F(p_i) = -\ln\left(\frac{1}{n_i!} (p_i N)^{n_i} \exp(-p_i N)\right) - \ln\left(\frac{1}{m_i!} (p_i M)^{m_i} \exp(-p_i M)\right)$$

$$= -n_i \ln(p_i) + p_i N + \ln(n_i!) - n_i \ln(N) - m_i \ln(p_i) + p_i M + \ln(m_i!) - m_i \ln(M)$$

minimieren:

$$\frac{dF}{dp_i} = -\frac{n_i}{p_i} + N - \frac{m_i}{p_i} + M \overset{!}{=} 0$$

$$\Leftrightarrow \quad (N+M)\, p_i = n_i + m_i$$

$$\Leftrightarrow \quad \hat{p}_i = \frac{n_i + m_i}{N + M}$$

(c) Stellen sie die χ^2-Test-Statistik unter Annahme der Nullhypothese auf. (Keine Vereinfachung des Terms nötig)

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - \langle \hat{n}_i \rangle)^2}{\widehat{Var}(n_i)} + \frac{(m_i - \langle \hat{m}_i \rangle)^2}{\widehat{Var}(m_i)}$$

$\langle \hat{n}_i \rangle \mathrel{\hat{=}}$ Schätzwert des Erwartungswert von n_i

$$\chi^2 = \sum \frac{(n_i - N\hat{p}_i)^2}{\ \ } + \frac{(m_i - M\hat{p}_i)^2}{\ \ }$$

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - N\hat{p}_i)^2}{N\hat{p}_i} + \frac{(m_i - M\hat{p}_i)^2}{M\hat{p}_i}$$

$$\text{mit } \hat{p}_i = \frac{n_i + m_i}{N+M}$$

(d) Wie viele Freiheitsgrade hat die X^2 Verteilung? Folgt die Test-Statistik für kleine Bininhalte $(n_i, m_i < 10)$ immer noch einer χ^2-Verteilung? Falls nein, warum nicht?

#Freiheitsgrade = # Messwerte − (# unabhängige Parameter)

unabhängige Parameter = # Parameter − # Bedingungen

hier:

#Messwerte = $2r$

#Parameter = $r + 2$ (wegen p_i, N und M)

#Bedingungen = 3 $\leftrightarrow \sum_{i=1}^{r} n_i = N, \sum_{i=1}^{r} m_i = M, \sum_{i=1}^{r} p_i = 1$

$\Rightarrow$ # Freiheitsgrade $= 2r - (r+2-3)$
$= r + 1$

Für kleine Bininhalte gilt nicht, dass die Poisson-Verteilung annähernd einer Gauß-Verteilung entspricht. Also folgt die Test-Statistik auch nicht einer χ^2-Verteilung.

(e) Gegeben sind die Histogramme:

n_1	n_2	n_3
111	188	333

m_1	m_2	m_3
15	36	30

Es kann gezeigt werden, dass die Test-Statistik zu $X^2 = \frac{1}{NM}\sum_{i=1}^{r}\frac{(Nm_i - Mn_i)^2}{n_i + m_i}$ vereinfachen lässt. Prüfen Sie, ob für die gegeben Histogramme die Nullhypothese für $\alpha = 0.1, 0.05, 0.01$ verworfen wird.

Werte einsetzen

$\Rightarrow \chi^2 \approx 8{,}429$

Freiheitsgrade $= r + 1 = 4$

Wenn der berechnete Wert größer ist, als der abgelesene Wert, wird die Nullhypothese abgelehnt.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x²								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58

als der abgelesene Wert, wird die Nullhypothese abgelehnt.

Werte ablesen:

12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57

$\alpha = 0.1$

$\Rightarrow 7{,}78 < 8{,}429$ $\Rightarrow$ Nullhypothese akzeptieren

$\alpha = 0.05$

$\Rightarrow 9{,}49 > 8{,}429$ $\Rightarrow$ Nullhypothese ablehnen

$\alpha = 0.01$

$\Rightarrow 13{,}28 > 8{,}429$ $\Rightarrow$ Nullhypothese ablehnen