

Kapitel 8

Signifikanzanalysen

8.1 Einführung

In den vorhergehenden Kapiteln haben wir Methoden kennengelernt, um aus Messungen Hypothesen abzuleiten, die mit den Daten verträglich sind. Es kann sich dabei um diskrete Hypothesen handeln oder auch um Funktionen, deren Parameter so bestimmt werden, dass die Funktion die beste Anpassung an die Daten darstellt. Die Bestimmung der Güte der Anpassung und der Signifikanz der Richtigkeit einer Hypothese haben wir für spezielle Fälle schon mehrfach angesprochen. Im Folgenden wollen wir allgemeiner statistische Tests zur Bestimmung der Signifikanz von Hypothesen besprechen, einerseits für die Signifikanz einer einzelnen Hypothese oder für die Entscheidung zwischen mehreren Hypothesen.

Wir nehmen an, es liegen Messwerte (x_1, \dots, x_n) vor, von denen eine Hypothese H_0 ('Null-Hypothese') abgeleitet wird, die zu prüfen ist. Zum Beispiel würde bei einer ML-Anpassung einer Funktion $f(x|\theta)$ die Funktion mit dem Parametersatz θ_0 , der die Likelihood-Funktion maximiert, der Null-Hypothese entsprechen. Zur Beurteilung der Signifikanz der Hypothese definiert man eine Testgröße t als eine Abbildung der Messdaten auf eine Größe, die möglichst die gesamte Information der Messung in komprimierter Form zusammenfasst:

$$(x_1, \dots, x_n) \rightarrow t(x_1, \dots, x_n | f, \theta_0) \quad (8.1)$$

Die Testgröße ('test statistic') hängt von den Messungen und der Hypothese H_0 ab, die hier durch die Anpassungsfunktion mit den Parametern θ_0 gegeben ist. Ein uns bereits bekanntes Beispiel für eine Testfunktion ist die χ^2 -Funktion. Die Testfunktion ist abhängig von der speziellen Stichprobe (x_1, \dots, x_n) und ist damit ebenfalls eine Zufallsvariable, die einer Wahrscheinlichkeitsverteilung $g(t)$ folgen soll. Dabei ist zu beachten, dass $g(t) = g(t|H_0)$ die Wahrscheinlichkeitsverteilung von t für eine feste Hypothese ist und damit von den Messwerten abhängt. Wir werden also keine Wahrscheinlichkeit für die Hypothese formulieren können, sondern nur die Wahrscheinlichkeit für die spezielle Messung bei einer gegebenen Hypothese erhalten.

Als Maß für das Vertrauen in eine Hypothese oder die Güte einer Parameteranpassung bilden wir den p-Wert:

$$p = \int_{t_{mess}}^{\infty} g(t|H_0) dt. \quad (8.2)$$

Der p-Wert (auch ‘Signifikanz’) ist also die Wahrscheinlichkeit bei Wiederholung der Messungen Ergebnisse zu erhalten, die so gut oder schlechter wie die betrachtete Messung mit der Hypothese verträglich sind. Eine Hypothese wird akzeptiert, wenn der p-Wert größer als ein vorgegebenes Signifikanzniveau α (gleich dem früher eingeführten Konfidenzniveau) ist. Man beachte, dass der p-Wert für eine bestimmte Messung bestimmt wird, während das Signifikanz- oder Vertrauensniveau eine vorgegebene Größe ist (zum Beispiel $\alpha = 5\%$ oder 10%). Weiterhin ist zu beachten, dass alle p-Werte gleich wahrscheinlich sind, wenn die Messungen tatsächlich den Verteilungen entsprechend der Hypothese folgen.

8.2 Prüfung von Hypothesen

In diesem Abschnitt sollen einige spezielle Hypothesentests behandelt werden.

8.2.1 χ^2 -Test

Der χ^2 -Test, der bereits in Abschnitt 4.3 besprochen wurde, wird benutzt, um Messwerte y_i , $i = 1, \dots, n$, an den Punkten x_i mit Erwartungswerten η_i zu vergleichen. Wenn $\eta_i = \eta(x_i|\theta_0)$ die Erwartungswerte von Verteilungen mit Varianzen σ_i^2 sind, ist die Testfunktion:

$$t = \chi^2 = \sum_{i=1}^n \frac{(y_i(x_i) - \eta_i)^2}{\sigma_i^2}. \quad (8.3)$$

Wenn die y_i Stichprobenwerte aus Normalverteilungen sind, folgt t einer χ^2 -Verteilung (4.23) mit $n_F = n - m$ Freiheitsgraden, wobei m die Anzahl der bestimmten Parameter ist. Der χ^2 -Test wird auch häufig für nur näherungsweise normalverteilte Messwerte benutzt. Ein häufig vorkommendes Beispiel ist die Beschreibung poisson-verteilter Histogrammeinträge n_i durch Erwartungswerte $\nu_i = \nu(i|\theta_0)$ mit Varianzen $\sigma_i^2 = \nu_i$ (also die Varianzen von den Erwartungswerten und nicht von den Messwerten abgeleitet):

$$t = \chi^2 = \sum_{i=1}^n \frac{(n_i(x_i) - \nu_i)^2}{\nu_i}. \quad (8.4)$$

Der p-Wert zu einem χ^2 -Wert χ_m^2 einer Messung mit n_F Freiheitsgraden ist in den Abbildungen 4.3 und 4.4 in Abschnitt 4.3 abzulesen.

8.2.2 Studentsche t-Verteilung

Die Fragestellung ist, ob der Mittelwert $\bar{x} = \sum_i x_i/n$ einer Stichprobe x_i , $i = 1, \dots, n$, mit einem theoretischen Mittelwert μ vereinbar ist. Die Varianz des Mittelwertes wird mit der Varianz der Stichprobe s^2 entsprechend (4.11) zu s^2/n abgeschätzt. Die Wahrscheinlichkeitsdichte für die Testgröße

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} \quad (8.5)$$

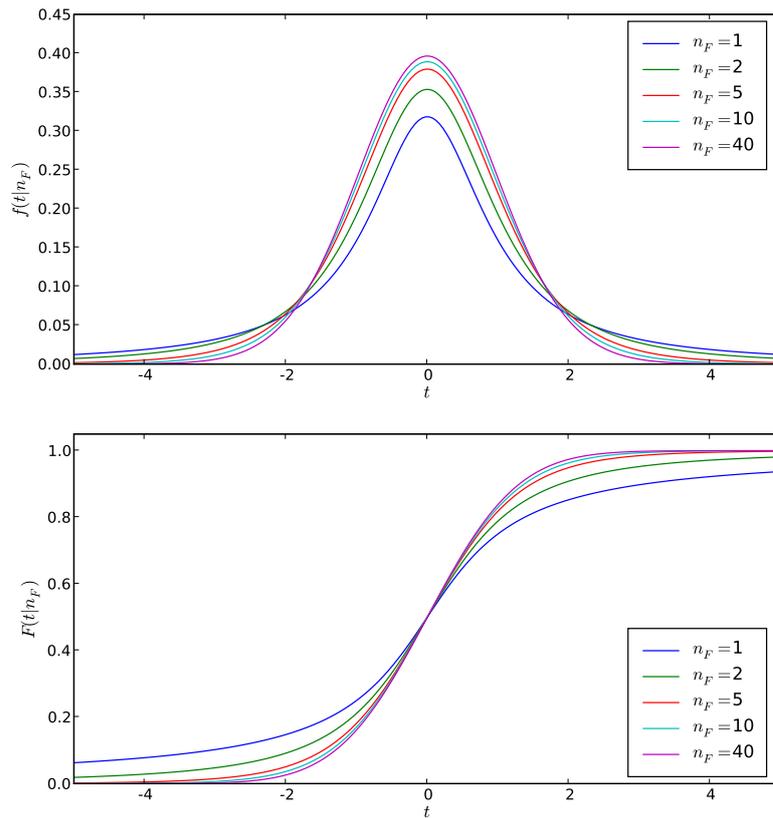


Abbildung 8.1: Oben: Die Studentsche t-Verteilung für verschiedene Freiheitsgrade k . Unten: kumulative Verteilungsfunktion der t-Verteilung.

folgt einer t-Verteilung,

$$f(t|n_F) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad (-\infty < t < +\infty). \quad (8.6)$$

Die Verteilung ist symmetrisch um $t = 0$, ist für n_F eine Cauchy-Verteilung $\sim 1/(1+t^2)$ und nähert sich für große n_F einer Gauss-Verteilung an (Abb. 8.1). Die t-Verteilung und deren kumulative Verteilungsfunktion findet man tabelliert in der entsprechenden Literatur. Das Python-Skript

```
from scipy import *
for t in [0.,0.5,1.0,1.5,2.0] :
    print t, stats.t.sf(t,10.)
```

berechnet folgende Tabelle für die p-Werte zu den angegebenen Werten von t und $n_F = 10$:

t	p [%]
0.0	50
0.5	31
1.0	17
1.5	8.2
2.0	3.7

Beispiel: Es seien drei Messungen ($x_1 = -1$, $x_2 = 0$, $x_3 = 1$) mit dem Mittelwert $\bar{x} = 0$ gegeben. Was ist der p-Wert, wenn der wahre Mittelwert $\mu = -1$ ist (Beispiel aus [3])? Mit den berechneten Zahlenwerten

$$s^2 = \frac{1}{2}(1 + 0 + 1) = 1.0, \quad t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \sqrt{3} = 1.732, \quad n_F = n - 1 = 2$$

ergibt das obige Python-Skript einen p-Wert von 11%. Das heißt, bei einem vorgegebenen Signifikanzniveau von zum Beispiel 5% oder 10% wäre die Hypothese zu akzeptieren.

8.2.3 F-Verteilung

Vergleich von Streuungen zweier Stichproben des Umfangs n_1 und n_2 mit gleichem Erwartungswert. Frage: haben beide Grundgesamtheiten die Gleiche Varianz. Die Fragestellung tritt zum Beispiel auf, wenn eine Größe mit zwei verschiedenen Apparaturen gemessen wird und zu klären ist, ob beide Apparaturen die gleiche Auflösung haben.

Dazu werden die empirischen Varianzen $s_1^2 = \chi_1^2/(n_1 - 1)$ und $s_2^2 = \chi_2^2/(n_2 - 1)$ nach (4.11) bestimmt. Die Testgröße ist der Quotient

$$F = \frac{s_1^2}{s_2^2} \tag{8.7}$$

Die Wahrscheinlichkeitsverteilung lässt sich aus mit Hilfe der χ^2 -Verteilungen zu den Freiheitsgraden $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ ableiten (Abb. 8.2), wenn die Stichproben normalverteilt sind:

$$f(F|\nu_1, \nu_2) = \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} \cdot \frac{\Gamma(\frac{\nu_1}{2} + \frac{\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \cdot \frac{F^{\frac{\nu_1}{2}-1}}{(\nu_1 F + \nu_2)^{\frac{\nu_1+\nu_2}{2}}} \quad (0 \leq F \leq +\infty) \tag{8.8}$$

Die Formel wird zum Beispiel in [1] abgeleitet. Der Erwartungswert der Verteilung ist

$$E(F) = \frac{\nu_2}{\nu_2 - 2} \quad \text{für } \nu_2 \gg 2. \tag{8.9}$$

Wegen des Quotienten in der Verteilung gilt

$$f(F_{12}|\nu_1, \nu_2) = f\left(\frac{1}{F_{12}}|\nu_2, \nu_1\right), \tag{8.10}$$

wobei jeweils ein F-Wert größer als 1 und der andere kleiner als 1 ist. Für einen Signifikanztest benutzt man üblicherweise den größeren der beiden Werte und verlangt wie auch bei den anderen Tests, dass das die Wahrscheinlichkeit, einen F-Wert größer

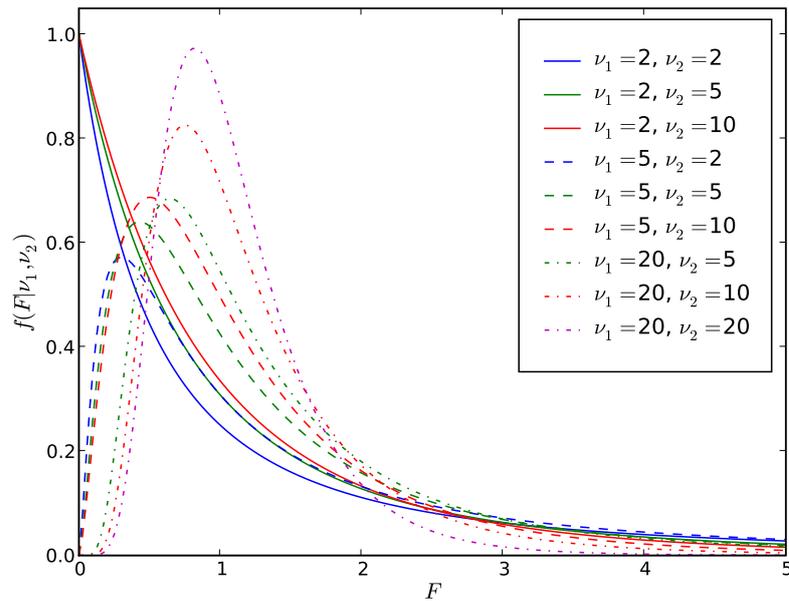


Abbildung 8.2: Wahrscheinlichkeitsdichte der F-Verteilung für verschiedene Freiheitsgrade der beiden beteiligten Stichproben.

als den gemessenen zu erhalten, ein vorgegebenes Signifikanzniveau übersteigt. Es ist allerdings zu beachten, dass mit der Einschränkung $F \geq 1$ die Normierung der F-Verteilung um eine Faktor 2 gegenüber der tabellierten Funktionen skaliert werden muss.

Man kann F-Werte und ihre Signifikanzen in Tabellen finden oder zum Beispiel mit Python berechnen:

```
>>> from scipy import *
>>> for F in [1.,2.,3.] : print F, 2.*stats.f.sf(F,10.,10.)
1.0 1.0
2.0 0.289691612051
3.0 0.0978546142578
```

Für vorgegebene Signifikanzniveaus kann man andererseits den dazugehörigen F-Wert berechnen:

```
>>> from scipy import *
>>> for p in [0.10,0.05,0.01] : print p, stats.f.isf(p/2.,10.,10.)
0.1 2.97823701608
0.05 3.7167918646
0.01 5.84667842506
```

Beispiel: Mit zwei Messapparaturen wird jeweils eine Messung gemacht. Die Ergebnisse sind: $n_1 = 10$, $s_1^2 = 3.7$; $n_2 = 7$, $s_2^2 = 6.5$ (aus [1]). Daraus ergibt sich $F = 6.5/3.7 = 1.8$ mit einem p-Wert von 41%, mit dem die Hypothese sicherlich akzeptiert wird.

```
>>> from scipy import *
>>> F=1.8
>>> print F, 2.*stats.f.sf(F,6.,9.)
1.8 0.410775850533
```

8.2.4 Kolmogorov-Smirnov-Test

Es geprüft werden, ob eine Stichprobe (x_1, \dots, x_n) einer Gesamtheit mit Wahrscheinlichkeitsdichte $f(x)$ entnommen ist. Dazu könnte man die Daten in x -Intervalle einteilen und mit einem χ^2 -Test die Hypothese überprüfen. Problematisch wird dieser Test bei kleinen Anzahlen in den Bins. Auch ist der χ^2 -Test nicht sehr sensitiv auf tendentielle Abweichungen nach oben oder unten in begrenzten x -Bereichen. Durch Einteilung der Daten in Überschuß- und Unterschussbereiche könnte man solche Tendenzen sichtbar machen. Aber wie bestimmt man dann die p-Werte, da ja eine solche Neueinteilung auf einer subjektiven Einschätzung beruht?

Mit dem Kolmogorov-Smirnov-Test kann man die Verträglichkeit der Stichprobe mit einer Wahrscheinlichkeitsdichte ohne Intervaleinteilung prüfen. Dazu wird die Verteilungsfunktion

$$F(x) = \int_{-\infty}^x f(\xi) d\xi \quad (8.11)$$

verglichen mit der Schätzung dieses Integrals mit Hilfe der Stichprobe:

$$F_n(x) = \frac{\text{Anzahl der } x_i\text{-Werte } \leq x}{n}. \quad (8.12)$$

Die Testgröße ist proportional zu der größten Differenz zwischen den beiden kumulativen Verteilungen:

$$t = \max |F_n(x) - F(x)|. \quad (8.13)$$

Werte von t zu vorgegebenen Signifikanzniveaus sind für verschiedene Freiheitsgrade in Tabelle 8.1 aufgelistet. Für große n_F -Werte ist der p-Wert durch eine unendliche Reihe gegeben:

$$p = 2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp(-2k^2 n_F t^2) \quad (8.14)$$

Mit dem Python-Skript

```
>>> from scipy import *
>>> t=0.447
>>> print t, stats.ksone.sf(t,5.)
0.447 0.099980005201
```

wird $p = 0.1$ für $t = 0.447$ bei $n_F = 5$ in der Tabelle 8.1 reproduziert. Andererseits lässt sich mit der inversen Funktion `stats.ksone.isf` bei vorgegebenem p-Wert oder Signifikanzniveau und Freiheitsgrad der dazugehörige t-Wert bestimmen:

```
>>> from scipy import *
>>> p=0.1
>>> print p, stats.ksone.isf(p,5.)
0.1 0.446980061221
```

Tabelle 8.1: Kolmogorov-Smirnov-Test (einseitig): Tabelle der Werte t_0 zu einem Signifikanzniveau für verschiedene Freiheitsgrade.

n_F	Signifikanzniveau				
	0.1	0.05	0.025	0.01	0.005
1	0.9000	0.9500	0.9750	0.9900	0.9950
2	0.6838	0.7764	0.8419	0.9000	0.9293
3	0.5648	0.6360	0.7076	0.7846	0.8290
4	0.4927	0.5652	0.6239	0.6889	0.7342
5	0.4470	0.5094	0.5633	0.6272	0.6685
6	0.4104	0.4680	0.5193	0.5774	0.6166
7	0.3815	0.4361	0.4834	0.5384	0.5758
8	0.3583	0.4096	0.4543	0.5065	0.5418
9	0.3391	0.3875	0.4300	0.4796	0.5133
10	0.3226	0.3687	0.4092	0.4566	0.4889
11	0.3083	0.3524	0.3912	0.4367	0.4677
12	0.2958	0.3382	0.3754	0.4192	0.4490
13	0.2847	0.3255	0.3614	0.4036	0.4325
14	0.2748	0.3142	0.3489	0.3897	0.4176
15	0.2659	0.3040	0.3376	0.3771	0.4042
16	0.2578	0.2947	0.3273	0.3657	0.3920
17	0.2504	0.2863	0.3180	0.3553	0.3809
18	0.2436	0.2785	0.3094	0.3457	0.3706
19	0.2373	0.2714	0.3014	0.3369	0.3612
20	0.2316	0.2647	0.2941	0.3287	0.3524
21	0.2262	0.2586	0.2872	0.3210	0.3443
22	0.2212	0.2528	0.2809	0.3139	0.3367
23	0.2165	0.2475	0.2749	0.3073	0.3295
24	0.2120	0.2424	0.2693	0.3010	0.3229
25	0.2079	0.2377	0.2640	0.2952	0.3166
26	0.2040	0.2332	0.2591	0.2896	0.3106
27	0.2003	0.2290	0.2544	0.2844	0.3050
28	0.1968	0.2250	0.2499	0.2794	0.2997
29	0.1935	0.2212	0.2457	0.2747	0.2947
30	0.1903	0.2176	0.2417	0.2702	0.2899
31	0.1873	0.2141	0.2379	0.2660	0.2853
32	0.1844	0.2108	0.2342	0.2619	0.2809
33	0.1817	0.2077	0.2308	0.2580	0.2768
34	0.1791	0.2047	0.2274	0.2543	0.2728
35	0.1766	0.2018	0.2242	0.2507	0.2690
36	0.1742	0.1991	0.2212	0.2473	0.2653
37	0.1719	0.1965	0.2183	0.2440	0.2618
38	0.1697	0.1939	0.2154	0.2409	0.2584
39	0.1675	0.1915	0.2127	0.2379	0.2552
40	0.1655	0.1891	0.2101	0.2349	0.2521
> 40	$1.07/\sqrt{n_F}$	$1.22/\sqrt{n_F}$	$1.36/\sqrt{n_F}$	$1.52/\sqrt{n_F}$	$1.63/\sqrt{n_F}$

Quelle: <http://www.york.ac.uk/depts/maths/tables>

8.3 Vertrauensintervalle

Die Angabe von Vertrauensintervallen im Parameterraum, das ist der Bereich in dem der gesuchte Satz von Parametern mit einer vorgegebenen Wahrscheinlichkeit liegt, ist problematisch, weil meistens die Wahrscheinlichkeiten für Parameter nicht bekannt sind. Deshalb entbrennen auf diesem Feld auch die heftigsten Kämpfe zwischen Bayes-Anhängern und Frequentisten. Im PDG Review [15] werden beide Sichtweisen diskutiert und weiterführende Literatur angegeben.

8.3.1 Bayes-Vertrauensintervalle

Die Wahrscheinlichkeitsdichte für die Parameter θ bei einem gegebenen Satz von Messung x ist nach dem Bayes-Theorem:

$$p(\theta|x) = \frac{L(x|\theta)p(\theta)}{\int L(x|\theta')p(\theta')d\theta'}. \quad (8.15)$$

Das Problem ist das die ‘A-Priori-Wahrscheinlichkeit’ $p(\theta)$ im allgemeinen nicht bekannt ist und Annahmen gemacht werden müssen (die einfachste Annahme wäre eine Gleichverteilung). Vorteilhaft ist diese Formulierung für den Ausschluss unphysikalischer Bereiche, in denen man $p(\theta) = 0$ setzen kann (zum Beispiel, damit eine Zählrate nicht negativ wird).

Das Intervall $[\theta_u, \theta_o]$, mit dem das gesuchte θ mit eine (Posterior-)Wahrscheinlichkeit von $1 - \alpha$ liegt, wird bestimmt zu:

$$1 - \alpha = \int_{\theta_u}^{\theta_o} p(\theta|x)d\theta \quad (8.16)$$

Das vorgegebene Vertrauensniveau $1 - \alpha$ kann mit verschiedenen Intervalgrenzen erreicht werden. Naheliegend ist eine Auswahl, so dass jeweils unterhalb und oberhalb des Intervalls die Wahrscheinlichkeiten $\alpha/2$ sind. Man kann auch den Vertrauensbereich so festzulegen, dass $p(\theta|x)$ in dem Bereich immer größer ist als außerhalb. Wenn man obere oder untere Ausschliessungsgrenzen zu einem Vertrauensniveau $1 - \alpha$ geben will, kann man in (8.16) $\theta_u = 0$ beziehungsweise $\theta_o = \infty$ setzen.

8.3.2 ‘Klassische’ Vertrauensintervalle

‘Frequentisten’ benutzen die Neyman-Konstruktion der Vertrauensintervalle wie in Abb. 8.3 gezeigt. Statt die Wahrscheinlichkeitsdichte für die Parameter bestimmt man die Wahrscheinlichkeitsdichte $f(x|\theta)$ der Messwerte x bei festen Parametern θ . Für verschiedene Parameter θ werden nun die Grenzen x_1 und x_2 bestimmt, in denen mit einer Wahrscheinlichkeit $1 - \alpha$ die Messwerte liegen:

$$P(x_1 < x < x_2|\theta) = 1 - \alpha = \int_{x_1}^{x_2} f(x|\theta)dx. \quad (8.17)$$

Diese Intervalle werden nun kontinuierlich als Funktion von θ bestimmt, so dass man das Band (‘confidence belt’) wie in Abb. 8.3 erhält. Diese Konstruktion kann, beziehungsweise sollte, vor der Messung gemacht werden. Wenn das Messergebnis

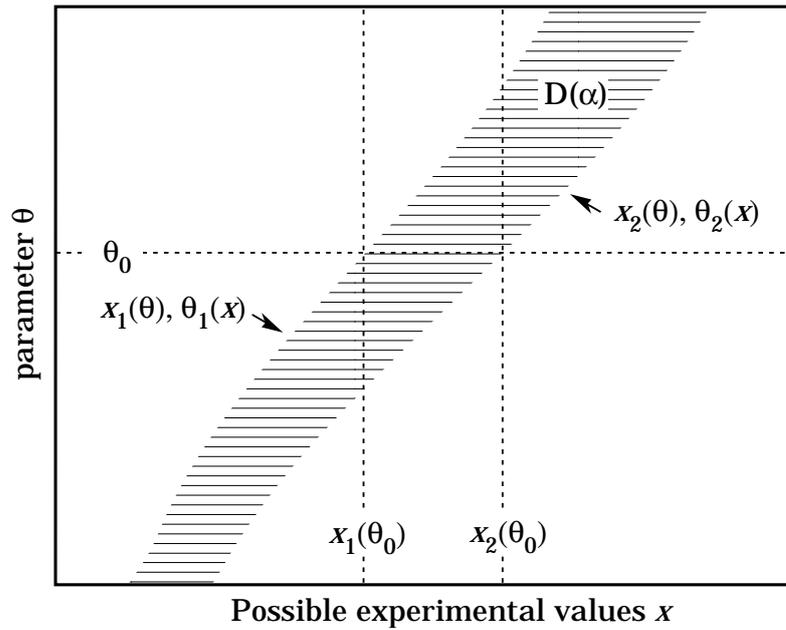


Abbildung 8.3: Konstruktion des Vertrauensbandes (siehe Text); aus [15].

dann x_0 ist, ergeben sich die unteren und oberen Grenzen θ_1, θ_2 als die Schnittpunkte der vertikalen Linie $x = x_0$ mit dem unteren beziehungsweise oberen Bandrand. Die Bandränder werden in Abb. 8.3 als Funktionen $\theta_1(x)$ und $\theta_2(x)$ bezeichnet.

Auch hier ist die Lage des Vertrauensintervalls zunächst nicht festgelegt. Feldmann und Cousins¹ haben eine Anordnung nach Likelihood-Verhältnissen vorgeschlagen. Bei der Bestimmung des Vertrauensintervalles x_1, x_2 zu festem θ (horizontal) wird zu jedem x -Wert der Parameter θ_{best} gesucht (entlang der Vertikalen), für den die Likelihood-Funktion an dieser x -Position maximal ist:

$$L(x|\theta_{best}) \geq L(x|\theta) \quad \forall \theta \text{ bei festem } x. \quad (8.18)$$

Das Verhältnis

$$\lambda = \frac{L(x|\theta)}{L(x|\theta_{best})} \quad (8.19)$$

wird als Funktion von x bei festem θ (also in der Horizontalen) bestimmt und es werden die x -Werte nach der Größe von λ geordnet, wobei dem Punkt mit dem größten λ der Rang 1 zugeordnet wird. Das Vertrauensintervall wird nun sukzessive durch Hinzunahme von x -Werten entsprechend ihrer Rangfolge so aufgebaut, bis das vorgegebene Vertrauensniveau $1 - \alpha$ erreicht ist. Dazu werden bei diskreten Verteilungen die Wahrscheinlichkeiten summiert und bei kontinuierlichen Verteilungen wird das entsprechende Integral in diskreten Schritten approximiert.

Die Feldmann-Cousins-Konstruktion stellt unter anderem sicher, dass die beste Parameteranpassung in dem Vertrauensintervall jedenfalls enthalten ist. Zudem liefert das Verfahren ein Rezept, wann als Ergebnis ein zentrales Vertrauensintervall und wann eine obere oder untere Grenze angegeben werden sollen. Eine Grenze wird angegeben, wenn das Band für eine x -Messung die untere oder obere Grenze des er-

¹G.J. Feldman and R.D. Cousins, Phys. Rev. D57, 3873 (1998).

Tabelle 8.2: Konstruktion der Vertrauensintervalle für ein Signal μ , wenn n Ereignisse gemessen werden und der Untergrund $b = 3.0$ ist. Das Beispiel in der Tabelle zeigt die Berechnung für $\mu = 0.5$.

n	$L(n \mu)$	μ_{best}	$L(n \mu_{best})$	λ	Rang
0	0.030	0.0	0.050	0.607	6
1	0.106	0.0	0.149	0.708	5
2	0.185	0.0	0.224	0.826	3
3	0.216	0.0	0.224	0.963	2
4	0.189	1.0	0.195	0.966	1
5	0.132	2.0	0.175	0.753	4
6	0.077	3.0	0.161	0.480	7
7	0.039	4.0	0.149	0.259	
8	0.017	5.0	0.140	0.121	
9	0.007	6.0	0.132	0.050	
10	0.002	7.0	0.125	0.018	
11	0.001	8.0	0.119	0.006	

laubten θ -Bereiches erreicht. Das ist am besten in folgendem Beispiel zu sehen, das aus der Veröffentlichung von Feldmann und Cousins stammt:

Beispiel: In einem Experiment soll eine bestimmte Reaktion untersucht werden. Als Kandidaten für die Reaktion werden n Ereignisse gezählt (die vorher benutzte Variable x ist jetzt also die diskrete Variable n), die sich aus einem Signalanteil s und einem Untergrundanteil b zusammensetzen. Der Erwartungswert des Untergrundes sei zu $b = 3.0$ bestimmt. Für verschiedene Messergebnisse n sollen 90%-Vertrauensintervalle für den Erwartungswert μ des Signals ermittelt werden. Die Rate n folgt einer Poisson-Verteilung,

$$L(n|\mu) = \frac{(\mu + b)^n}{n!} e^{-(\mu+b)} \quad (8.20)$$

Für die Konstruktion des Vertrauensbandes nimmt man sich in diskreten Schritten jeweils einen festen Wert $\mu \geq 0$ vor. Dann bildet man zu jedem möglichen Messergebnis n das Verhältnis

$$\lambda = \frac{L(n|\mu)}{L(n|\mu_{best})}, \quad (8.21)$$

wobei μ_{best} die beste μ -Schätzung für dieses n ist. Als Beispiel ist in Tab. 8.2 für $\mu = 0.5$ die Bestimmung der Likelihood-Ordnung gezeigt. Um ein 90%-Intervall zu erhalten addiert man die Wahrscheinlichkeiten der Ränge 1 bis 7, entsprechend $n = 0 - 6$, was 93.5% ergibt. Da die Summe bis Rang 6 nur 85.8% ergibt, entscheidet man sich für die 'konservativere' Lösung.

Wenn man diese Prozedur für den gesamten abzudeckenden μ -Bereich wiederholt hat, erhält man schliesslich die Darstellung des Vertrauensbandes in Abb. 8.4. Bei gemessenen Raten bis $n = 4$ wird das Band durch $\mu = 0$ begrenzt; deshalb würde man bei einem Messergebnis $n \leq 4$ eine obere Grenze für das Signal angeben.

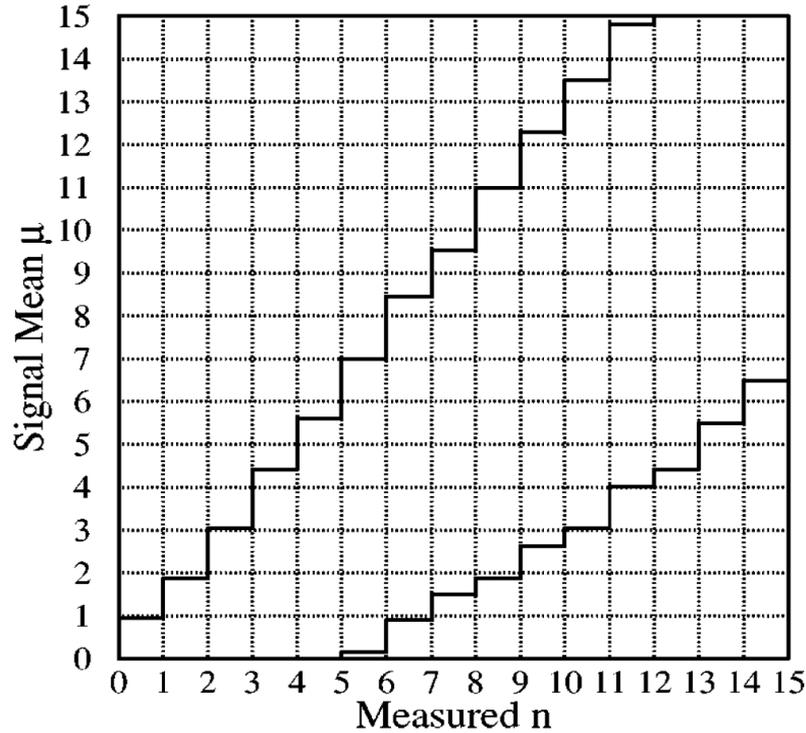


Abbildung 8.4: Vertrauensband zu 90% Vertrauensniveau für die Bestimmung einer Signalrate μ bei einem bekannten Untergrund von $b = 3.0$ (nach Feldman-Cousins).

Sensitivität: Experimentell bestimmte Ausschließungsgrenzen können wegen statistischer Fluktuationen, bei ansonsten gleichen Bedingungen, für verschiedene Experimente unterschiedlich ausfallen. Zur Beurteilung der Leistungsfähigkeit eines Experimentes ist es üblich, die ‘Sensitivität’ eines Experimentes auf eine Messgröße anzugeben, indem man die entsprechenden Vertrauensintervalle oder Grenzen für die Erwartungswerte angibt.

Beispiel: In dem obigen Beispiel ist die Sensitivität für die Hypothese $\mu = 0$ bei einem Untergrund $b = 3.0$ mit 90% Vertrauensniveau durch den Erwartungswert $E(\mu_{90\%}) = \langle \mu_{90\%} \rangle$ gegeben. Wenn man den Erwartungswert von $\mu_{90\%}$ durch

$$\langle \mu_{90\%} \rangle \approx \mu_{90\%}(E(n)|\mu = 0, b = 3) = \mu_{90\%}(n = 3) \quad (8.22)$$

nähert, entnimmt man aus der Abb. 8.4 eine obere Grenze von etwa $\langle \mu_{90\%} \rangle = 4.3$.

Bei einem Beschleunigerexperiment muss man $\langle \mu_{90\%} \rangle$ durch die integrierte Luminosität L und die Akzeptanz ϵ dividieren, um den Wirkungsquerschnitt zu erhalten, den man im Mittel mit 90% Vertrauensniveau ausschliessen kann:

$$\langle \sigma_{90\%} \rangle = \frac{\langle \mu_{90\%} \rangle}{\epsilon L}. \quad (8.23)$$

Bei kosmischer Strahlung muss man entsprechend durch die effektive Detektorfläche A und die Zeitspanne der Datennahme T teilen, um die Sensitivität

auf einen Teilchenfluß zu bestimmen:

$$\langle \phi_{90\%} \rangle = \frac{\langle \mu_{90\%} \rangle}{AT}. \quad (8.24)$$