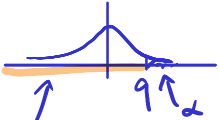
ein Stichproben-t-Test

$\Theta = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$

Modell: n u.i.v. Beobachtungen $X_1, \ldots, X_n$, $X_i \sim \mathcal{N}_{\mu,\sigma^2}$, $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt.

$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$, $S^2 := \frac{1}{n-1} \sum_{i=1}^{n-1} (X_i - \overline{X})^2$

Wir wissen: Für jedes $\mu_0 \in \mathbb{R}$, $\sigma^2 > 0$ ist unter $P_{(\mu_0,\sigma^2)}$



$$T := \sqrt{n}\frac{\overline{X} - \mu_0}{\sqrt{S^2}} \sim \text{Student-}(n-1)$$

(vgl. Satz 2.5)

$= \{(\mu_0,\sigma^2) : \sigma > 0\}$

Signifikanzniveau $\alpha \in (0, 1)$:

Zweiseitiger Test von $H_0 : \{\mu = \mu_0\}$ gegen $H_1 : \{\mu \neq \mu_0\}$:
Lehne H_0 ab, wenn $|T| > q_{n-1,1-\alpha/2}$, wobei $q_{n-1,1-\alpha/2}$ = $(1 - \alpha/2)$-Quantil der Student-$(n - 1)$-Vert.

$\{(\mu,\sigma^2) : \sigma > 0, \mu \neq \mu_0\}$

Einseitiger Test von $H_0 : \{\mu \leq \mu_0\}$ gegen $H_1 : \{\mu > \mu_0\}$:
Lehne H_0 ab, wenn $T > q_{n-1,1-\alpha}$, wobei $q_{n-1,1-\alpha}$ = $(1 - \alpha)$-Quantil der Student-$(n - 1)$-Vert.

Beispiel:
Die Wirksamkeit eines gewissen Schlafmittels soll geprüft werden.

10 Patienten erhalten das Schlafmittel, die Anzahl zusätzlicher Stunden Schlaf wird in einer Nacht beobachtet.

Wir nehmen an, die Beob. sind u.i.v. $\sim \mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu = 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Die Daten*

Patient i	1	2	3	4	5	6	7	8	9	10
zus. Schl.	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0

Es ist $n = 10$, $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 0.75$, $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 \approx 1.79$,
$t = \frac{\overline{x}-0}{s/\sqrt{10}} \approx 1.326$

Das 0.975-Quantil der Student-9-Verteilung ist ≈ 2.262, demnach können wir die Nullhypothese nicht ablehnen.

(Für ein Student-9-verteiltes T ist $P(|T| \geq 1.326) \approx 0.2176$, dies ist der p-Wert des Tests.)

* Aus Student (William S. Gosset), The Probable Error of a Mean, Biometrika 6:1–25 (1908)

Man kann unseren Befund folgendermaßen formulieren:

„Die Beobachtungen sind mit der Nullhypothese $\mu = 0$ (im statistischen Sinne) verträglich."

oder

„Die beobachtete Abweichung $\overline{x} = 0.75$ ist nicht signifikant von 0 verschieden (t-Test, $\alpha = 0.05$)."

Das Beispiel in R:

```
> schlaf <- c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4,
              3.7, 0.8, 0.0, 2.0)
> t.test(schlaf)

One Sample t-test

data:  schlaf
t = 1.3257, df = 9, p-value = 0.2176
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.5297804  2.0297804
sample estimates:
mean of x
     0.75
```

Beispiel:
Die Wirksamkeit eines Schlafmittels soll mit der eines anderen verglichen werden

10 Patienten erhalten Schlafmittel A, die Anzahl zusätzlicher Stunden Schlaf wird in einer Nacht beobachtet.
Dann erhalten dieselben 10 Patienten Schlafmittel B, wieder wird die Anzahl zusätzlicher Stunden Schlaf wird in einer Nacht beobachtet.

Da dieselben Patienten untersucht werden, können (und sollten) wir die Messungen paaren:
Wir interessieren uns bei jedem Patienten für die Differenz des (zusätzlichen) Schlafs bei Mittel B und bei Mittel A.

Wir nehmen an, die beobachteten Differenzen sind Realisierungen von u.i.v. ZVn mit Vert. $\mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu \leq 0$ gegen die Alternative $\mu > 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Wir nehmen an, die beobachteten Differenzen sind Realisierungen von u.i.v. ZVn mit Vert. $\mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu \leq 0$ gegen die Alternative $\mu > 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Dies wäre beispielsweise in folgender Situation angemessen: Wir möchten darlegen, dass Mittel *B* wirksamer ist als Mittel *A*, indem wir die Nullhypothese „$\mu \leq 0$“ entkräften.

Wir nehmen an, die beobachteten Differenzen sind Realisierungen von u.i.v. ZVn mit Vert. $\mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu \leq 0$ gegen die Alternative $\mu > 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Dies wäre beispielsweise in folgender Situation angemessen: Wir möchten darlegen, dass Mittel *B* wirksamer ist als Mittel *A*, indem wir die Nullhypothese „$\mu \leq 0$“ entkräften.

Die Daten*

Patient *i*	1	2	3	4	5	6	7	8	9	10
Mittel *A*	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
Mittel *B*	1.9	0.8	1.1	0.1	-0.1	4.4	5.5	1.6	4.6	3.4
Diff.	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

* Aus Student (William S. Gosset), The Probable Error of a Mean, Biometrika 6:1–25 (1908)

Patient i	1	2	3	4	5	6	7	8	9	10
Diff.	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

Es ist $n = 10$, $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1.58$, $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 \approx 1.23$, $t = \frac{\overline{x}-0}{s/\sqrt{10}} \approx 4.062$

Das 0.95-Quantil der Student-9-Verteilung ist ≈ 1.833, demnach können wir die Nullhypothese ablehnen.

(Für ein Student-9-verteiltes T ist $P(T > 4.062) \approx 0.0014$, dies ist der p-Wert des Tests.)

Patient i	1	2	3	4	5	6	7	8	9	10
Diff.	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

Es ist $n = 10$, $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1.58$, $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 \approx 1.23$,
$t = \frac{\overline{x}-0}{s/\sqrt{10}} \approx 4.062$

Das 0.95-Quantil der Student-9-Verteilung ist ≈ 1.833,
demnach können wir die Nullhypothese ablehnen.

(Für ein Student-9-verteiltes T ist $P(T > 4.062) \approx 0.0014$, dies ist der p-Wert des Tests.)

Mögliche knappe Formulierung:

„Die beobachtete Differenz $\overline{x} = 1.58$ ist signifikant größer als 0 (einseitiger t-Test, $\alpha = 0.05$).“

Das Beispiel in R:

```
> diff <- c(1.2, 2.4, 1.3, 1.3, 0.0, 1.0, 1.8,
            0.8, 4.6, 1.4)
> t.test(diff, alternative="greater")

One Sample t-test

data:  diff
t = 4.0621, df = 9, p-value = 0.001416
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 0.8669947       Inf
sample estimates:
mean of x
     1.58
```

Inhalt

Schätzer und Konfidenzintervall
Punktschätzer
Konfidenzintervall, bekannte Varianz
Konfidenzintervall, unbekannte Varianz

Statistische Tests
Abstrakter Rahmen
ein Stichproben- oder gepaarter t-Test
zwei Stichproben-t-Test mit Annahme gleicher Varianzen
Bericht: Welchs zwei Stichproben-t-Test

ungepaarter *t*-Test: Modell

m u.i.v. Beobachtungen $X_1, \ldots, X_m$ und davon unabhängig n u.i.v. Beobachtungen $Y_1, \ldots, Y_n$, unter P_ϑ

$X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, $Y_j \sim \mathcal{N}_{\mu_2,\sigma^2}$ mit $\vartheta = (\mu_1, \mu_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$

Seien
$$\overline{X} := \frac{1}{m} \sum_{i=1}^{m} X_i, \ \overline{Y} := \frac{1}{n} \sum_{j=1}^{n} Y_j$$

$$S_X^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \overline{X})^2, \ S_Y^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_j - \overline{Y})^2,$$

ungepaarter *t*-Test: Modell

m u.i.v. Beobachtungen $X_1, \dots, X_m$ und davon unabhängig n u.i.v. Beobachtungen $Y_1, \dots, Y_n$, unter P_ϑ

$X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, $Y_j \sim \mathcal{N}_{\mu_2,\sigma^2}$ mit $\vartheta = (\mu_1, \mu_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$

Seien
$$\overline{X} := \frac{1}{m}\sum_{i=1}^{m} X_i, \ \overline{Y} := \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2, \ S_Y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2,$$

$$S^2 := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} \left(= \frac{1}{m+n-2}\Big(\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2 \Big) \right),$$

(bem.: $\mathbb{E}_{(\mu_1,\mu_2,\sigma^2)}[S^2] = \sigma^2$)

ungepaarter t-Test: Modell

m u.i.v. Beobachtungen $X_1, \dots, X_m$ und davon unabhängig n u.i.v. Beobachtungen $Y_1, \dots, Y_n$, unter P_ϑ
$X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, $Y_j \sim \mathcal{N}_{\mu_2,\sigma^2}$ mit $\vartheta = (\mu_1, \mu_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$

Seien

$$\overline{X} := \frac{1}{m}\sum_{i=1}^{m} X_i,\ \overline{Y} := \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2,\ S_Y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2,$$

$$S^2 := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} \left(= \frac{1}{m+n-2}\Big(\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2\Big)\right),$$

(bem.: $\mathbb{E}_{(\mu_1,\mu_2,\sigma^2)}[S^2] = \sigma^2$)

$$T = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}}.$$

$\mathrm{Var}[\overline{X} - \overline{Y}]$
$= \mathrm{Var}[\overline{X}] + \mathrm{Var}[\overline{Y}]$
$= \frac{1}{m^2} \cdot m \cdot \sigma^2 + \frac{1}{n^2} \cdot n \cdot \sigma^2 = \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)$

Für $\mu_0 \in \mathbb{R}$, $\sigma^2 > 0$ ist unter $P_{(\mu_0,\mu_0,\sigma^2)}$ (d.h. $\mu_1 = \mu_2 = \mu_0$

$$T = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim \text{Student-}(m + n - 2)$$

Signifikanzniveau $\alpha \in (0, 1)$:

Zweiseitiger Test von $H_0 : \{\mu_1 = \mu_2\}$ gegen $H_1 : \{\mu_1 \neq \mu_2\}$:
Lehne H_0 ab, wenn $|T| > q_{m+n-2,1-\alpha/2}$, wobei $q_{m+n-2,1-\alpha/2}$ = $(1 - \alpha/2)$-Quantil der Student-$(m + n - 2)$-Vert.

Einseitiger Test von $H_0 : \{\mu_1 \leq \mu_2\}$ gegen $H_1 : \{\mu_1 > \mu_2\}$:
Lehne H_0 ab, wenn $T > q_{m+n-2,1-\alpha}$, wobei $q_{m+n-2,1-\alpha}$ = $(1 - \alpha)$-Quantil der Student-$(m + n - 2)$-Vert.

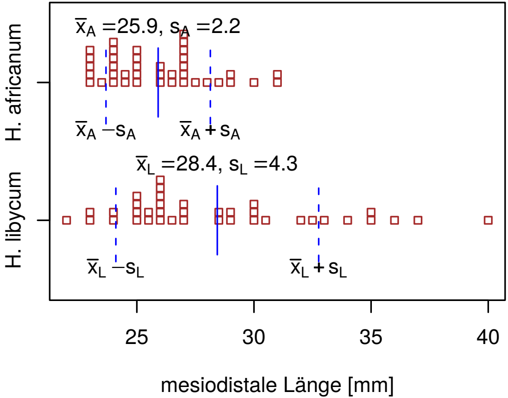
Beispiel: Es wurden fossile Backenzähne gefunden, die zwei Arten von Urpferden zugeordnet wurden, und jeweils die („mesiodistale“) Länge bestimmt.

Wir möchten die (Null-)Hypothese prüfen, ob die mittlere Zahnlänge bei den beiden Arten gleich ist.

Beispiel: Es wurden fossile Backenzähne gefunden, die zwei Arten von Urpferden zugeordnet wurden, und jeweils die („mesiodistale“) Länge bestimmt.

Wir möchten die (Null-)Hypothese prüfen, ob die mittlere Zahnlänge bei den beiden Arten gleich ist.

Die Daten



Hipparion africanum

$n_A = 39$

$\overline{x}_A = 25.9$

$s_A = 2.2$

Hipparion lybicum

$n_L = 38$

$\overline{x}_L = 28.4$

$s_L = 4.3$

Wir verwenden Signifikanzniveau $\alpha = 0.01$, das 99,5%-Quantil der Student-Vert. mit 75 Freiheitsgraden ist ≈ 2.64.
Es ist

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_L - 1)s_L^2}{n_A + n_L - 1} \approx 11.94, \quad t = \frac{\overline{x}_A - \overline{x}_L}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_L}}} \approx -3.229,$$

Wir können die Nullhypothese „die mittlere mesiodistale Länge bei *H. lybicum* und bei *H. africanum* sind gleich“ zum Signifikanzniveau 1% ablehnen.

(Für ein Student-75-verteiltes T ist $P(|T| > 3.229) \approx 0.0018$, dies ist der p-Wert des Tests.)

Wir verwenden Signifikanzniveau $\alpha = 0.01$, das 99,5%-Quantil der Student-Vert. mit 75 Freiheitsgraden ist ≈ 2.64.
Es ist

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_L - 1)s_L^2}{n_A + n_L - 1} \approx 11.94, \quad t = \frac{\overline{x}_A - \overline{x}_L}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_L}}} \approx -3.229,$$

Wir können die Nullhypothese „die mittlere mesiodistale Länge bei *H. lybicum* und bei *H. africanum* sind gleich" zum Signifikanzniveau 1% ablehnen.

(Für ein Student-75-verteiltes T ist $P(|T| > 3.229) \approx 0.0018$, dies ist der p-Wert des Tests.)

Mögliche Formulierung unseres Befunds:

„Die mittlere mesiodistale Länge war signifikant größer (28.4 mm) bei *H. libycum* als bei *H. africanum* (25.9 mm) (t-Test, $\alpha = 0{,}01$)."

Das Beispiel in R:

```
> t.test(md[Art=="africanum"],md[Art=="libycum"],
         var.equal=TRUE)

Two Sample t-test

data:  md[Art == "africanum"] and md[Art == "libycu
t = -3.2289, df = 75, p-value = 0.001845
alternative hypothesis:
true difference in means is not equal to 0
95 percent confidence interval:
 -4.0811448 -0.9667634
sample estimates:
mean of x mean of y
 25.91026  28.43421
```

Betrachten wir (spaßeshalber) nochmal die Schlafmittel-Daten, diesmal ungepaart:

```
> medA <- c(0.7,-1.6,-0.2,-1.2,-0.1,3.4,3.7,0.8,0.0,2.0)
> medB <- c(1.9,0.8,1.1,0.1,-0.1,4.4,5.5,1.6,4.6,3.4)

> t.test(medB, medA, var.equal=TRUE)

Two Sample t-test

data:  medB and medA
t = 1.8608, df = 18, p-value = 0.07919
alternative hypothesis: true difference in means is
95 percent confidence interval:
 -0.203874  3.363874
sample estimates:
mean of x mean of y
     2.33      0.75
```

Inhalt

Schätzer und Konfidenzintervall
Punktschätzer
Konfidenzintervall, bekannte Varianz
Konfidenzintervall, unbekannte Varianz

Statistische Tests
Abstrakter Rahmen
ein Stichproben- oder gepaarter t-Test
zwei Stichproben-t-Test mit Annahme gleicher Varianzen
Bericht: Welchs zwei Stichproben-t-Test

t-Statistik ohne Annahme gleicher Varianz

Es gibt auch eine Version des zwei-Stichproben-t-Tests, der die Annahme gleicher Varianzen nicht trifft (wir werden ihn im Verlauf der Vorlesung allerdings nicht verwenden):

Wäre eine beobachtete Abweichung $\overline{x} - \overline{y}$ mit der Nullhypothese verträglich, dass $\mu_X = \mu_Y$?

t-Statistik ohne Annahme gleicher Varianz

Es gibt auch eine Version des zwei-Stichproben-t-Tests, der die Annahme gleicher Varianzen nicht trifft (wir werden ihn im Verlauf der Vorlesung allerdings nicht verwenden):

Wäre eine beobachtete Abweichung $\overline{x} - \overline{y}$ mit der Nullhypothese verträglich, dass $\mu_X = \mu_Y$?

Wir schätzen die Streuung von $\overline{X} - \overline{Y}$ durch

$$\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} \quad \text{und bilden} \quad T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}}.$$

Unter $P_{(\mu_0,\mu_0,\sigma_1^2,\sigma_2^2)}$ ist T „approximativ Student-verteilt mit g Freiheitsgraden“

wobei g aus den Daten geschätzt wird, $g = \dfrac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{s_X^4}{n_X^2(n_X-1)} + \frac{s_Y^4}{n_Y^2(n_Y-1)}}$

Welchs* zwei Stichproben-t-Test

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}}, \qquad g = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{s_X^4}{n_X^2(n_X-1)} + \frac{s_Y^4}{n_Y^2(n_Y-1)}}$$

Man verwirft die Nullhypothese „$\mu_1 = \mu_2$" (zum Niveau α), wenn

$$\texttt{pt}(|t|, \texttt{df=}g, \texttt{lower.tail=FALSE}) \leq \alpha/2$$

ist, d.h. wenn die Wahrscheinlichkeit, dass eine Student-verteilte Zufallsgröße mit g Freiheitsgraden einen betragsmäßig mindestens so großen Wert wie den beobachteten t-Wert annimmt, $\leq \alpha$ ist.
(Analoges Vorgehen für einseitige Tests)

---

* B. L. Welch, The Significance of the Difference between Two Means When the Population Variances Are Unequal, Biometrika 29:350–362, (1938)

Zwei-Stichproben-*t*-Test mit R:

```
> A <- md[Art=="africanum"]
> L <- md[Art=="libycum"]
> t.test(L,A)

Welch Two Sample t-test

data:  L and A
t = 3.2043, df = 54.975, p-value = 0.002255
alternative hypothesis: true difference in means
is not equal to 0
95 percent confidence interval:
 0.9453745 4.1025338
sample estimates:
mean of x mean of y
 28.43421  25.91026
```

Statistik für Informatiker, SS 2021

2. Ideen aus der Statistik

2.3 Schätzprinzipien

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/StatInfo21/

Wir haben bisher (nur) den empirischen Mittelwert

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

als Schätzer für den Erwartungswert einer (unbekannten) Verteilung verwendet.

Das ist sehr naheliegend und diese Schätzer haben auch „gute“ Eigenschaften (Erwartungstreue, Konsistenz, asymptotische Normalität), wie wir in Kapitel 2.2 gesehen haben.

Wir haben bisher (nur) den empirischen Mittelwert

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

als Schätzer für den Erwartungswert einer (unbekannten) Verteilung verwendet.

Das ist sehr naheliegend und diese Schätzer haben auch „gute“ Eigenschaften (Erwartungstreue, Konsistenz, asymptotische Normalität), wie wir in Kapitel 2.2 gesehen haben.

Andererseits trifft man auch Situationen, in denen zumindest auf den ersten Blick kein „offensichtlicher“ Schätzer auf der Hand liegt; (spätestens) dann lohnen sich systematischere Ansätze, von denen wir einige in diesem Kapitel betrachten.

Inhalt

ML-Schätzer
Beispiele

Bayes-Statistik
Beispiel: Münzwurf mit zufälliger Erfolgswahrscheinlichkeit

Kleinste-Quadrate-Schätzer und lineare Regression
Beispiel: Größen von Vätern und Söhnen

Schätzer, allgemein

Wir betrachten eine Zufallsvariable X mit Wertebereich S in einem statistischen Modell $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta \in \Theta})$.

Wir verwenden X, um die Beobachtungen zu modellieren (oft schreiben wir $X = (X_1, X_2, \dots, X_n)$, wenn das Experiment aus n Wiederholungen oder „Bauteilen“ besteht).

Schätzer, allgemein

Wir betrachten eine Zufallsvariable X mit Wertebereich S in einem statistischen Modell $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta \in \Theta})$.

Wir verwenden X, um die Beobachtungen zu modellieren (oft schreiben wir $X = (X_1, X_2, \dots, X_n)$, wenn das Experiment aus n Wiederholungen oder „Bauteilen“ besteht).

Abstrakt gesehen ist ein Schätzer für ϑ eine Funktion $T : S \to \Theta$ (mit der Interpretation, dass wir anhand von Beobachtungen $x \in S$ im Rahmen des Modells „tippen“ würden, dass der Wert von ϑ wohl $T(x)$ ist).

Definition 2.7

Man nennt diese Situation ein *statistisches Standardmodell*, wenn entweder S diskret ist oder $S \subset \mathbb{R}^n$ gilt und X unter P_ϑ

$$\text{Gewichte } \rho_\vartheta(\cdot) \quad \text{bzw.} \quad \text{Dichte } \rho_\vartheta(\cdot) \qquad \text{für } \vartheta \in \Theta$$

besitzt.

Definition 2.7

Man nennt diese Situation ein *statistisches Standardmodell*, wenn entweder S diskret ist oder $S \subset \mathbb{R}^n$ gilt und X unter P_ϑ

$$\text{Gewichte } \rho_\vartheta(\cdot) \quad \text{bzw.} \quad \text{Dichte } \rho_\vartheta(\cdot) \qquad \text{für } \vartheta \in \Theta$$

besitzt. Die Funktion

$$\begin{array}{rll} \rho: & S \times \Theta & \to [0, \infty) \\ & \Psi & \\ & (x, \vartheta) & \mapsto \rho(x, \vartheta) := \rho_\vartheta(x) \end{array}$$

heißt die *Likelihood-Funktion* (manchmal auch „Plausibilitäts-Funktion")

Definition 2.7

Man nennt diese Situation ein *statistisches Standardmodell*, wenn entweder S diskret ist oder $S \subset \mathbb{R}^n$ gilt und X unter P_ϑ

$$\text{Gewichte } \rho_\vartheta(\cdot) \quad \text{bzw.} \quad \text{Dichte } \rho_\vartheta(\cdot) \qquad \text{für } \vartheta \in \Theta$$

besitzt. Die Funktion

$$\begin{aligned} \rho : \quad & S \times \Theta && \to [0, \infty) \\ & \quad \Psi && \\ & (x, \vartheta) && \mapsto \rho(x, \vartheta) := \rho_\vartheta(x) \end{aligned}$$

heißt die *Likelihood-Funktion* (manchmal auch „Plausibilitäts-Funktion"), für $x \in S$ heißt

$$L_x : \Theta \to [0, \infty), \quad L_x(\vartheta) = \rho(x, \vartheta)$$

die Likelihood-Funktion zum Beobachtungswert x.

ML-Schätzer

Definition 2.8

Ein Schätzer $T : S \to \Theta$ heißt (ein) *Maximum-Likelihood-Schätzer*, wenn

$$\rho(x, T(x)) = \max_{\vartheta \in \Theta} \rho(x, \vartheta) \quad \forall x \in S$$

(auch kurz ML-Schätzer genannt, engl. MLE = maximum likelihood estimator).

Man schreibt oft auch $\widehat{\vartheta}_{\mathrm{ML}}$ für den ML-Schätzer.

Inhalt

ML-Schätzer
Beispiele

Bayes-Statistik
Beispiel: Münzwurf mit zufälliger Erfolgswahrscheinlichkeit

Kleinste-Quadrate-Schätzer und lineare Regression
Beispiel: Größen von Vätern und Söhnen

Beispiel 2.9

1. („Rückfangmethode“, engl. “capture-recapture”)

Ein Teich enthalte ϑ Fische (einer gewissen Art, $\vartheta \in \mathbb{N}$ ist der unbekannte Parameter), fange und markiere m, setze wieder aus. Wenn sich die markierten Fische gut verteilt haben, fange erneut n Fische.

Beispiel 2.9

1. („Rückfangmethode“, engl. “capture-recapture”)

Ein Teich enthalte ϑ Fische (einer gewissen Art, $\vartheta \in \mathbb{N}$ ist der unbekannte Parameter), fange und markiere m, setze wieder aus. Wenn sich die markierten Fische gut verteilt haben, fange erneut n Fische.

Nehmen wir an, wir beobachten unter den erneut gefangenen x markierte Fische.

Beispiel 2.9

1. („Rückfangmethode“, engl. “capture-recapture”)

Formalisierung als statistisches Modell: $S = \{0, 1, \dots, n\}$, $\Theta = \{(m \vee n), (m \vee n) + 1, (m \vee n) + 2, \dots\}$, unter P_ϑ ist die beobachtete Anzahl $X \sim \mathrm{Hyp}_{m,\vartheta-m,n}$ (hypergeometrische Verteilung, s.a. Bsp. 1.15).

Beispiel 2.9

1. („Rückfangmethode“, engl. “capture-recapture”)

Formalisierung als statistisches Modell: $S = \{0, 1, \dots, n\}$, $\Theta = \{(m \vee n), (m \vee n) + 1, (m \vee n) + 2, \dots\}$, unter P_ϑ ist die beobachtete Anzahl $X \sim \mathrm{Hyp}_{m,\vartheta-m,n}$ (hypergeometrische Verteilung, s.a. Bsp. 1.15).

Die Likelihood-Funktion ist demnach

$$\rho(x, \vartheta) = \frac{\binom{m}{x}\binom{\vartheta-m}{n-x}}{\binom{\vartheta}{n}}$$

Beispiel 2.9

1. („Rückfangmethode“, engl. “capture-recapture”)

Formalisierung als statistisches Modell: $S = \{0, 1, \dots, n\}$, $\Theta = \{(m \vee n), (m \vee n) + 1, (m \vee n) + 2, \dots\}$, unter P_ϑ ist die beobachtete Anzahl $X \sim \mathrm{Hyp}_{m,\vartheta-m,n}$ (hypergeometrische Verteilung, s.a. Bsp. 1.15).

Die Likelihood-Funktion ist demnach

$$\rho(x, \vartheta) = \frac{\binom{m}{x}\binom{\vartheta-m}{n-x}}{\binom{\vartheta}{n}},$$

der ML-Schätzer ist

$$\widehat{\vartheta}_{\mathrm{ML}} = \left\lfloor \frac{n}{x} \cdot m \right\rfloor,$$

(mit $\lfloor a \rfloor = \max\{z \in \mathbb{Z} : z \le a\}$).

Beispiel 2.9

1. („Rückfangmethode“, engl. “capture-recapture”)

Formalisierung als statistisches Modell: $S = \{0, 1, \dots, n\}$, $\Theta = \{(m \vee n), (m \vee n) + 1, (m \vee n) + 2, \dots\}$, unter P_ϑ ist die beobachtete Anzahl $X \sim \mathrm{Hyp}_{m,\vartheta-m,n}$ (hypergeometrische Verteilung, s.a. Bsp. 1.15).

Die Likelihood-Funktion ist demnach

$$\rho(x, \vartheta) = \frac{\binom{m}{x}\binom{\vartheta-m}{n-x}}{\binom{\vartheta}{n}},$$

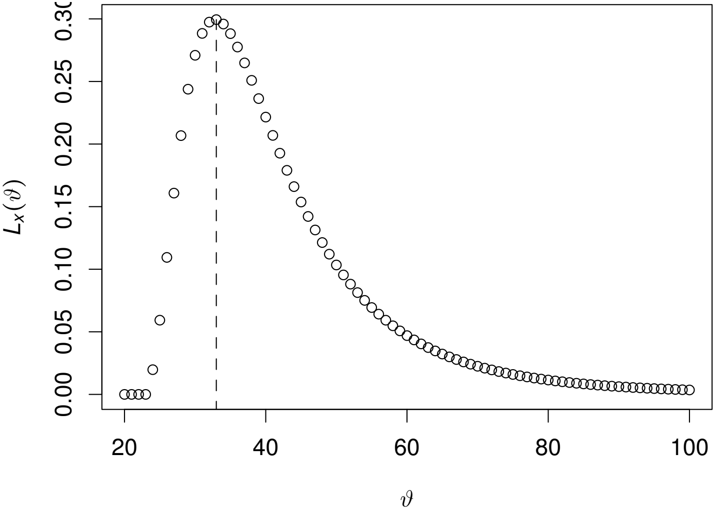
der ML-Schätzer ist

$$\widehat{\vartheta}_{\mathrm{ML}} = \left\lfloor \frac{n}{x} \cdot m \right\rfloor,$$

(mit $\lfloor a \rfloor = \max\{z \in \mathbb{Z} : z \le a\}$).

Dies ist auch anschaulich plausibel, denn $\dfrac{m}{\widehat{\vartheta}_{\mathrm{ML}}} \approx \dfrac{x}{n}$.

Likelihood-Funktion $L_x(\vartheta)$, hier
$m = 10, n = 20, x = 6$ ($\widehat{\vartheta}_{\mathrm{ML}} = 33$)



$\widehat{\vartheta}_{\mathrm{ML}} = \left\lfloor \frac{n}{x} \cdot m \right\rfloor$: Es ist nämlich

$$\frac{\rho(x,\vartheta)}{\rho(x,\vartheta-1)} = \frac{\binom{\vartheta-m}{n-x}\binom{\vartheta-1}{n}}{\binom{\vartheta}{n}\binom{\vartheta-1-m}{n-x}}$$

$$= \frac{(\vartheta-m)(\vartheta-n)}{\vartheta(\vartheta-m-n+x)} = 1 - \frac{\vartheta x - mn}{\vartheta(\vartheta-m-n+x)} \begin{cases} > 1, & \vartheta < \frac{mn}{x}, \\ = 1, & \vartheta = \frac{mn}{x}, \\ < 1, & \vartheta > \frac{mn}{x} \end{cases}$$

(beachte: stets ist $\vartheta - m \geq n - x$, es gibt im Teich mindestens so viele unmarkierte Fische wie in der Rückfang-Stichprobe).

$\widehat{\vartheta}_{\mathrm{ML}} = \left\lfloor \frac{n}{x} \cdot m \right\rfloor$: Es ist nämlich

$$\frac{\rho(x,\vartheta)}{\rho(x,\vartheta-1)} = \frac{\binom{\vartheta-m}{n-x}\binom{\vartheta-1}{n}}{\binom{\vartheta}{n}\binom{\vartheta-1-m}{n-x}}$$

$$= \frac{(\vartheta-m)(\vartheta-n)}{\vartheta(\vartheta-m-n+x)} = 1 - \frac{\vartheta x - mn}{\vartheta(\vartheta-m-n+x)} \begin{cases} >1, & \vartheta < \frac{mn}{x}, \\ =1, & \vartheta = \frac{mn}{x}, \\ <1, & \vartheta > \frac{mn}{x} \end{cases}$$

(beachte: stets ist $\vartheta - m \geq n - x$, es gibt im Teich mindestens so viele unmarkierte Fische wie in der Rückfang-Stichprobe).

Bem.: Falls $\frac{mn}{x} \in \mathbb{N}$, so ist der ML-Schätzer hier nicht eindeutig: $\frac{mn}{x} - 1$ und $\frac{mn}{x}$ maximieren die Likelihood.

Beispiel 2.9

2. (Erfolgsw'keit im Binomialmodell per ML schätzen)

Ein zufälliges Experiment mit zwei möglichen Ausgängen (Idealisierung: Werfen einer Münze) werde unabhängig (unter identischen Bedingungen) n-fach wiederholt, wir zählen die Anzahl X der „Erfolge".

$S = \{0, 1, \dots, n\}$, unter P_ϑ, $\vartheta \in \Theta = [0, 1]$ ist $X \sim \mathrm{Bin}_{n,\vartheta}$, also

$$\rho(x, \vartheta) = L_x(\vartheta) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}$$

Beispiel 2.9

2. (Erfolgsw'keit im Binomialmodell per ML schätzen)

$\rho(x, \vartheta) = L_x(\vartheta) = \binom{n}{x}\vartheta^x(1 - \vartheta)^{n-x}$

Beispiel 2.9

2. (Erfolgsw'keit im Binomialmodell per ML schätzen)
$\rho(x, \vartheta) = L_x(\vartheta) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x}$

$$\begin{aligned} \frac{d}{d\vartheta} \log L_x(\vartheta) &= \frac{d}{d\vartheta}\Big(\log \binom{n}{x} + x \log \vartheta + (n-x) \log(1-\vartheta) \Big) \\ &= \frac{x}{\vartheta} - \frac{n-x}{1-\vartheta} = 0 \iff \vartheta = \frac{x}{n} \end{aligned}$$

Beispiel 2.9

2. (Erfolgsw'keit im Binomialmodell per ML schätzen)
$\rho(x, \vartheta) = L_x(\vartheta) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}$

$$\begin{aligned} \frac{d}{d\vartheta} \log L_x(\vartheta) &= \frac{d}{d\vartheta}\Big(\log \binom{n}{x} + x \log \vartheta + (n - x) \log(1 - \vartheta)\Big) \\ &= \frac{x}{\vartheta} - \frac{n - x}{1 - \vartheta} = 0 \iff \vartheta = \frac{x}{n} \end{aligned}$$

d.h. hier ist $\widehat{\vartheta}_{\mathrm{ML}} = \dfrac{x}{n}$.
(Es ist $\frac{d}{d\vartheta} \log L_x(\vartheta) > 0$ für $\vartheta < x/n$ und $\frac{d}{d\vartheta} \log L_x(\vartheta) < 0$ für $\vartheta > x/n$, d.h. es handelt sich tatsächlich um ein Maximum; Inspektion zeigt, dass auch in den Randfällen $x = 0$ und $x = n$ $\widehat{\vartheta}_{\mathrm{ML}} = \frac{x}{n}$ gilt.)

Beispiel 2.9

2. (Erfolgsw'keit im Binomialmodell per ML schätzen)
$\rho(x, \vartheta) = L_x(\vartheta) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$

$$\frac{d}{d\vartheta}\log L_x(\vartheta) = \frac{d}{d\vartheta}\Big(\log\binom{n}{x} + x\log\vartheta + (n-x)\log(1-\vartheta)\Big)$$
$$= \frac{x}{\vartheta} - \frac{n-x}{1-\vartheta} = 0 \iff \vartheta = \frac{x}{n}$$

d.h. hier ist $\widehat{\vartheta}_{\mathrm{ML}} = \dfrac{x}{n}$.
(Es ist $\frac{d}{d\vartheta}\log L_x(\vartheta) > 0$ für $\vartheta < x/n$ und $\frac{d}{d\vartheta}\log L_x(\vartheta) < 0$ für $\vartheta > x/n$, d.h. es handelt sich tatsächlich um ein Maximum; Inspektion zeigt, dass auch in den Randfällen $x = 0$ und $x = n$ $\widehat{\vartheta}_{\mathrm{ML}} = \frac{x}{n}$ gilt.)

$\widehat{\vartheta}_{\mathrm{ML}}$ ist hier auch ein sehr „naheliegender" Schätzer: Wir schätzen die (unbekannte) Erfolgswahrscheinlichkeit durch die relative Anzahl der beobachteten Erfolge.

Beispiel 2.9

3. (Normales Modell mit bekannter Varianz)

n Beobachtungen seien u.i.v. $\sim \mathcal{N}_{\vartheta,\sigma^2}$, $\sigma^2 > 0$ sei bekannt und $\vartheta \in \Theta = \mathbb{R}$ soll geschätzt werden.

Mit $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ ist

$$\begin{aligned} \rho(x, \vartheta) = L_x(\vartheta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x_i - \vartheta)^2}{2\sigma^2} \Big) \\ &= (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \vartheta)^2 \Big) \end{aligned}$$

d.h.

$$L_x(\vartheta) \stackrel{!}{=} \max \iff \sum_{i=1}^{n} (x_i - \vartheta)^2 \stackrel{!}{=} \min .$$

Beispiel 2.9

3. (Normales Modell mit bekannter Varianz)

$$L_x(\vartheta) \stackrel{!}{=} \max \iff \sum_{i=1}^{n}(x_i - \vartheta)^2 \stackrel{!}{=} \min .$$

Mit $\overline{x} := \dfrac{1}{n}\sum_{i=1}^{n} x_i$ ist

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \vartheta)^2 = \left(\overline{x} - \vartheta\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2,$$

d.h. es ist $\widehat{\vartheta}_{\mathrm{ML}} = \overline{x}$, das empirische Mittel der Beobachtungen.

(Für die Gleichung beachte $\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2$.)

Beispiel 2.9

3. (Normales Modell mit bekannter Varianz)

Die ML-Schätzer $\widehat{\vartheta}_{\mathrm{ML}} = \overline{x}$ stimmen hier mit dem naheliegenden Lageschätzer aus Kap. 2.2 überein, sie sind insbesondere konsistent und (hier sogar exakt) normalverteilt.

(Dies gilt asymptotisch für $n \to \infty$ recht allgemein für ML-Schätzer.)

Beispiel 2.9

4. (Normales Modell, unbekannter Erwartungswert und unbekannte Varianz)

n Beobachtungen seien u.i.v. $\sim \mathcal{N}_{\mu,v}$ mit unbekannten $\mu \in \mathbb{R}$ und $v \in (0, \infty)$.

(Formalisierung: $S = \mathbb{R}^n$, $\Theta = \{\vartheta = (\mu, v) : \mu \in \mathbb{R}, v > 0\}$, unter $P_{(\mu,v)}$ ist $X = (X_1, \ldots, X_n) \sim \mathcal{N}_{\mu,v}^{\otimes n}$)

Beispiel 2.9

4. (Normales Modell, unbekannter Erwartungswert und unbekannte Varianz)

n Beobachtungen seien u.i.v. $\sim \mathcal{N}_{\mu,v}$ mit unbekannten $\mu \in \mathbb{R}$ und $v \in (0,\infty)$.

(Formalisierung: $S = \mathbb{R}^n$, $\Theta = \{\vartheta = (\mu, v) : \mu \in \mathbb{R}, v > 0\}$, unter $P_{(\mu,v)}$ ist $X = (X_1, \dots, X_n) \sim \mathcal{N}_{\mu,v}^{\otimes n}$)

Wie in 3. ist

$$\log L_x\big((\mu, v)\big) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log v - \frac{1}{2v}\sum_{i=1}^{n}(x_i - \mu)^2,$$

nach obigem ist $\widehat{\mu}_{\mathrm{ML}} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ Maximierer bezüglich μ (für jeden Wert von v).

Beispiel 2.9

4. (Normales Modell, unbekannter Erwartungswert und unbekannte Varianz)

$$\log L_x\big((\mu, v)\big) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log v - \frac{1}{2v}\sum_{i=1}^{n}(x_i - \mu)^2,$$

somit ist

$$\frac{\partial}{\partial v}\log L_x\big((\mu, v)\big)\Big|_{\mu=\widehat{\mu}_{\mathrm{ML}}} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathrm{ML}})^2$$

Beispiel 2.9

4. (Normales Modell, unbekannter Erwartungswert und unbekannte Varianz)

$$\log L_x\big((\mu, v)\big) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log v - \frac{1}{2v}\sum_{i=1}^{n}(x_i - \mu)^2,$$

somit ist

$$\frac{\partial}{\partial v}\log L_x\big((\mu, v)\big)\Big|_{\mu=\widehat{\mu}_{\mathrm{ML}}} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathrm{ML}})^2$$

also $\dfrac{\partial}{\partial v}\log L_x\big((\widehat{\mu}_{\mathrm{ML}}, v)\big) = 0 \iff v = \widehat{v}_{\mathrm{ML}} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathrm{ML}})^2$.

(Und man prüft: $\log L_x\big((\widehat{\mu}_{\mathrm{ML}}, v)\big)$ ist wachsend für $v < \widehat{v}_{\mathrm{ML}}$, fallend für $v > \widehat{v}_{\mathrm{ML}}$.)

Beispiel 2.9

4. (Normales Modell, unbekannter Erwartungswert und unbekannte Varianz)

$$\log L_x\big((\mu, v)\big) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log v - \frac{1}{2v}\sum_{i=1}^{n}(x_i - \mu)^2,$$

somit ist

$$\frac{\partial}{\partial v}\log L_x\big((\mu, v)\big)\Big|_{\mu=\widehat{\mu}_{\mathrm{ML}}} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathrm{ML}})^2$$

also $\dfrac{\partial}{\partial v}\log L_x\big((\widehat{\mu}_{\mathrm{ML}}, v)\big) = 0 \iff v = \widehat{v}_{\mathrm{ML}} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathrm{ML}})^2$.

(Und man prüft: $\log L_x\big((\widehat{\mu}_{\mathrm{ML}}, v)\big)$ ist wachsend für $v < \widehat{v}_{\mathrm{ML}}$, fallend für $v > \widehat{v}_{\mathrm{ML}}$.)

Beachte: Der ML-Schätzer für die unbekannte Varianz ist hier die (unkorrigierte) Stichprobenvarianz, also ist er nicht erwartungstreu (vgl. Kap. 2.2; allerdings ist er für halbwegs großes *n* auch nicht weit weg von Erwartungstreue).

Beispiel 2.9

5. Beobachtungen seien u.i.v. uniform auf $\{1, 2, \ldots, \vartheta\}$ (mit einem unbekannten $\vartheta \in \mathbb{N}$).

Beispiel 2.9

5. Beobachtungen seien u.i.v. uniform auf $\{1, 2, \ldots, \vartheta\}$ (mit einem unbekannten $\vartheta \in \mathbb{N}$).

Mögliche Interpretation: In einer Stadt gibt es ϑ Taxis, die mit $1, 2, \ldots, \vartheta$ durchnummeriert sind. Wir beobachten die Nummern von *n* zufällig gewählten Taxis und schätzen, wieviele Taxis es insgesamt gibt.

Beispiel 2.9

5. Beobachtungen seien u.i.v. uniform auf $\{1, 2, \ldots, \vartheta\}$ (mit einem unbekannten $\vartheta \in \mathbb{N}$).

Mögliche Interpretation: In einer Stadt gibt es ϑ Taxis, die mit $1, 2, \ldots, \vartheta$ durchnummeriert sind. Wir beobachten die Nummern von n zufällig gewählten Taxis und schätzen, wieviele Taxis es insgesamt gibt.

Formalisierung: $X = (X_1, \ldots, X_n)$ mit Werten in $S = \mathbb{N}^n$, $\rho\big((x_1, \ldots, x_n), \vartheta\big) = \vartheta^{-n} \mathbf{1}_{x_1, \ldots, x_n \leq \vartheta}$ für $\vartheta \in \Theta = \mathbb{N}$.

Beispiel 2.9

5. Beobachtungen seien u.i.v. uniform auf $\{1, 2, \ldots, \vartheta\}$ (mit einem unbekannten $\vartheta \in \mathbb{N}$).

Mögliche Interpretation: In einer Stadt gibt es ϑ Taxis, die mit $1, 2, \ldots, \vartheta$ durchnummeriert sind. Wir beobachten die Nummern von n zufällig gewählten Taxis und schätzen, wieviele Taxis es insgesamt gibt.

Formalisierung: $X = (X_1, \ldots, X_n)$ mit Werten in $S = \mathbb{N}^n$, $\rho\big((x_1, \ldots, x_n), \vartheta\big) = \vartheta^{-n}\mathbf{1}_{x_1,\ldots,x_n \leq \vartheta}$ für $\vartheta \in \Theta = \mathbb{N}$.

Es ist

$$\widehat{\vartheta}_{\mathrm{ML}} = \max\{x_1, x_2, \ldots, x_n\},$$

Beispiel 2.9

5. Beobachtungen seien u.i.v. uniform auf $\{1, 2, \dots, \vartheta\}$ (mit einem unbekannten $\vartheta \in \mathbb{N}$).

Mögliche Interpretation: In einer Stadt gibt es ϑ Taxis, die mit $1, 2, \dots, \vartheta$ durchnummeriert sind. Wir beobachten die Nummern von n zufällig gewählten Taxis und schätzen, wieviele Taxis es insgesamt gibt.

Formalisierung: $X = (X_1, \dots, X_n)$ mit Werten in $S = \mathbb{N}^n$, $\rho\big((x_1, \dots, x_n), \vartheta\big) = \vartheta^{-n} \mathbf{1}_{x_1,\dots,x_n \le \vartheta}$ für $\vartheta \in \Theta = \mathbb{N}$.

Es ist

$$\widehat{\vartheta}_{\mathrm{ML}} = \max\{x_1, x_2, \dots, x_n\},$$

denn

$$L_{(x_1,\dots,x_n)}(\vartheta) = \begin{cases} \dfrac{1}{\vartheta^n}, & \text{falls } \vartheta \ge x_1, x_2, \dots, x_n, \\ 0, & \text{sonst.} \end{cases}$$

Beispiel 2.9

6. n Beobachtungen seien u.i.v. $\sim \mathrm{Poi}_\vartheta$
(mit einem unbekannten $\vartheta \in \Theta := (0, \infty)$), d.h.
$\rho\big((x_1, \ldots, x_n), \vartheta\big) = \prod_{i=1}^n e^{-\vartheta} \frac{\vartheta^{x_i}}{x_i!}$

Beispiel 2.9

6. n Beobachtungen seien u.i.v. $\sim \mathrm{Poi}_\vartheta$
(mit einem unbekannten $\vartheta \in \Theta := (0, \infty)$), d.h.
$\rho\big((x_1, \ldots, x_n), \vartheta\big) = \prod_{i=1}^n e^{-\vartheta} \frac{\vartheta^{x_i}}{x_i!}$
Für $x = (x_1, \ldots, x_n)$ ist

$$L_x(\vartheta) = \frac{\exp\big(- n\vartheta + (x_1 + \cdots + x_n) \log \vartheta\big)}{x_1! x_2! \cdots x_n!}$$

Beispiel 2.9

6. n Beobachtungen seien u.i.v. $\sim \mathrm{Poi}_\vartheta$
(mit einem unbekannten $\vartheta \in \Theta := (0, \infty)$), d.h.
$\rho\big((x_1, \dots, x_n), \vartheta\big) = \prod_{i=1}^n e^{-\vartheta} \frac{\vartheta^{x_i}}{x_i!}$
Für $x = (x_1, \dots, x_n)$ ist

$$L_x(\vartheta) = \frac{\exp\big(- n\vartheta + (x_1 + \cdots + x_n) \log \vartheta\big)}{x_1! x_2! \cdots x_n!}$$

also

$$\frac{\partial}{\partial \vartheta} \log L_x(\vartheta) = -n + \frac{x_1 + \cdots + x_n}{\vartheta} \stackrel{!}{=} 0 \quad \Longleftrightarrow \quad \vartheta = \widehat{\vartheta}_{\mathrm{ML}}$$

mit $\widehat{\vartheta}_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^n x_i$, d.h. der ML-Schätzer ist hier wiederum der empirische Mittelwert.

Beispiel 2.9

6. n Beobachtungen seien u.i.v. $\sim \mathrm{Poi}_\vartheta$
(mit einem unbekannten $\vartheta \in \Theta := (0, \infty)$), d.h.
$\rho\big((x_1, \ldots, x_n), \vartheta\big) = \prod_{i=1}^n e^{-\vartheta} \frac{\vartheta^{x_i}}{x_i!}$
Für $x = (x_1, \ldots, x_n)$ ist

$$L_x(\vartheta) = \frac{\exp\big(-n\vartheta + (x_1 + \cdots + x_n)\log\vartheta\big)}{x_1!x_2!\cdots x_n!}$$

also

$$\frac{\partial}{\partial\vartheta}\log L_x(\vartheta) = -n + \frac{x_1 + \cdots + x_n}{\vartheta} \stackrel{!}{=} 0 \quad\Longleftrightarrow\quad \vartheta = \widehat{\vartheta}_{\mathrm{ML}}$$

mit $\widehat{\vartheta}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^n x_i$, d.h. der ML-Schätzer ist hier wiederum der empirische Mittelwert.

(Wegen $\mathbb{E}[X] = \vartheta$ für $X \sim \mathrm{Poi}_\vartheta$ ist es auch zugleich der sogenannte „Momentenschätzer“.)

Betrachen wir Beispiel 1.21 nochmals in diesem Licht:

Ladislaus von Bortkewitsch, *Das Gesetz der kleinen Zahlen*, Teubner, 1898, § 12, 4. („Die durch Schlag eines Pferdes im preußischen Heere getöteten") berichtet für 20 Jahre (1875–1894) und 10 Armeekops der preußischen Kavallerie, also insgesamt $20 \cdot 10 = 200$ „Korpsjahre" berichtet, in wievielen davon sich x Todesfälle durch Schlag eines Pferds ereigneten (Tabelle b) auf S. 25):

Ergebnis x	Anz. „Korpsjahre"	$200 \times \mathrm{Poi}_{0,61}(x)$
0	109	108,67
1	65	66,29
2	22	20,22
3	3	4,11
4	1	0,63
≥ 5	0	0,08

Ergebnis x	Anz. „Korpsjahre“	$200 \times \mathrm{Poi}_{0,61}(x)$
0	109	108,67
1	65	66,29
2	22	20,22
3	3	4,11
4	1	0,63
≥ 5	0	0,08

Sei X_i = Anzahl der im i-ten Korpsjahr Getöteten $\sim \mathrm{Poi}_\vartheta$, u.a. für verschiedene i.

Ergebnis x	Anz. „Korpsjahre“	$200 \times \mathrm{Poi}_{0,61}(x)$
0	109	108,67
1	65	66,29
2	22	20,22
3	3	4,11
4	1	0,63
≥5	0	0,08

Sei X_i = Anzahl der im i-ten Korpsjahr Getöteten $\sim \mathrm{Poi}_\vartheta$, u.a. für verschiedene i.

Die Daten geben uns zwar nicht die Beobachtungen x_i selbst, enthalten aber genug Informationen, um

$$\begin{aligned}\widehat{\vartheta}_{\mathrm{ML}} &= \frac{1}{200}\sum_{i=1}^{200} x_i \\ &= \frac{109}{200}\cdot 0 + \frac{65}{200}\cdot 1 + \frac{22}{200}\cdot 2 + \frac{3}{200}\cdot 3 + \frac{1}{200}\cdot 4 + 0 = 0{,}61\end{aligned}$$

zu berechnen.

Bericht („gute“ Eigenschaften der ML-Schätzer)**.**

Betrachte ein Produktmodell, d.h. $X = (X_1, \dots, X_n)$ mit Werten in $S = S_1^n$ und

$$\rho\big((x_1, \dots, x_n), \vartheta\big) = \prod_{i=1}^{n} \rho^{(1)}(x_i, \vartheta)$$

für eine Likelihood-Funktion $\rho^{(1)}$ auf $S_1 \times \Theta$ (unter P_ϑ sind die Beobachtungen $X_1, \dots, X_n$ u.i.v. mit Gewichten / Dichte $\rho^{(1)}(\cdot, \vartheta)$).

Bericht 2.10 („gute" Eigenschaften der ML-Schätzer)

Sei $\widehat{\vartheta}_{\mathrm{ML},n} = \widehat{\vartheta}_{\mathrm{ML},n}(X_1, \ldots, X_n)$ der ML-Schätzer basierend auf n unabhhängigen Beobachtungen, dann gilt (unter gewissen Bedingungen an $\rho^{(1)}(\cdot,\cdot)$) :

- Die ML-Schätzer sind konsistent, d.h. $P_\vartheta\big(|\widehat{\vartheta}_{\mathrm{ML},n} - \vartheta| \geq \varepsilon\big) \underset{n\to\infty}{\longrightarrow} 0$ für jedes $\varepsilon > 0$ und jedes $\vartheta \in \Theta$.
- Die ML-Schätzer sind asymptotisch normal mit (asymptotisch optimaler) Varianz $1/(nI(\vartheta))$, d.h. $\sqrt{n}\big(\widehat{\vartheta}_{\mathrm{ML},n} - \vartheta\big) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}_{0,1/I(\vartheta)}$ unter P_ϑ für jedes $\vartheta \in \Theta$, wobei $I(\vartheta) := \mathrm{Var}_\vartheta\Big[\dfrac{d}{d\vartheta} \log \rho(X, \vartheta)\Big]$ die sogenannte Fisher-Information ist.

Bericht 2.10 („gute“ Eigenschaften der ML-Schätzer)

Sei $\widehat{\vartheta}_{\mathrm{ML},n} = \widehat{\vartheta}_{\mathrm{ML},n}(X_1, \dots, X_n)$ der ML-Schätzer basierend auf n unabhhängigen Beobachtungen, dann gilt (unter gewissen Bedingungen an $\rho^{(1)}(\cdot,\cdot)$) :

- Die ML-Schätzer sind konsistent, d.h. $P_\vartheta\big(|\widehat{\vartheta}_{\mathrm{ML},n} - \vartheta| \geq \varepsilon\big) \underset{n\to\infty}{\longrightarrow} 0$ für jedes $\varepsilon > 0$ und jedes $\vartheta \in \Theta$.
- Die ML-Schätzer sind asymptotisch normal mit (asymptotisch optimaler) Varianz $1/(nI(\vartheta))$, d.h. $\sqrt{n}\big(\widehat{\vartheta}_{\mathrm{ML},n} - \vartheta\big) \overset{d}{\underset{n\to\infty}{\longrightarrow}} \mathcal{N}_{0,1/I(\vartheta)}$ unter P_ϑ für jedes $\vartheta \in \Theta$, wobei $I(\vartheta) := \mathrm{Var}_\vartheta\Big[\frac{d}{d\vartheta}\log\rho(X,\vartheta)\Big]$ die sogenannte Fisher-Information ist.

(Bem.: Im Fall des ML-Schätzers für den Erwartungswert im normalen Modell haben wir dies schon gesehen;
die Botschaft hier ist: es gilt ziemlich allgemein.)

Inhalt

ML-Schätzer
Beispiele

Bayes-Statistik
Beispiel: Münzwurf mit zufälliger Erfolgswahrscheinlichkeit

Kleinste-Quadrate-Schätzer und lineare Regression
Beispiel: Größen von Vätern und Söhnen

Wir folgten bisher (und werden dies auch in den späteren Kapiteln wieder tun) dem klassischen, sogenannten frequentistischen Ansatz der Statistik:

Wir fassen eine Menge Θ von „Parametern" ins Auge, für $\Theta \ni \vartheta$ beschreibt (in einem statistischen Modell) P_ϑ die Verteilung der Beobachungen, wenn dieses ϑ der tatsächlich zutreffende (sozusagen der „wahre") Parameter ist. In der konkreten Anwendungssituation kennen wir dieses „wahre" ϑ natürlich i.A. nicht, wir fassen es als zwar unbekannte, aber prinzipiell feste Größe auf. Wahrscheinlichkeitsaussagen beziehen sich *nicht* auf ϑ, sondern auf zufällige Beobachtungen unter P_ϑ.

Ansatz der Bayes-Statistik

Dies ist anders in der *Bayes-Statistik*: Man wählt eine Wahrscheinlichkeitsverteilung α auf Θ, die *a priori*-Verteilung (auch *Vorbewertung*) und stellt sich vor, dass die Daten einem zweistufigen Experiment entstammen:

- Zunächst wird der Parameter ϑ gemäß der a priori-Verteilung α erzeugt (insbesondere ist ϑ jetzt selbst eine Zufallsvariable),
- dann werden die Beobachtungen X zufällig erzeugt mit einer Verteilung, die vom gewählten ϑ abhängt.

Insbesondere besitzt in dieser Formulierung das Paar (X, ϑ) eine *gemeinsame Verteilung*.

Es gelte: Die a priori-Verteilung auf Θ hat Dichte bzw. Gewichte $\alpha(\vartheta)$

(je nachdem, ob ϑ kontinuierlich oder diskret verteilt ist; wir betrachten im Folgenden nur den Fall, dass $\Theta \subset \mathbb{R}$ ein Intervall ist und ϑ eine Dichte besitzt)

Interpretation: Ohne Kenntnis der Beobachtungen nehmen wir an, dass ϑ die a priori-Verteilung besitzt (z.B. aus „Erfahrung“ oder aus „Expertenwissen“).

Es gelte: Die a priori-Verteilung auf Θ hat Dichte bzw. Gewichte $\alpha(\vartheta)$

(je nachdem, ob ϑ kontinuierlich oder diskret verteilt ist; wir betrachten im Folgenden nur den Fall, dass $\Theta \subset \mathbb{R}$ ein Intervall ist und ϑ eine Dichte besitzt)

Interpretation: Ohne Kenntnis der Beobachtungen nehmen wir an, dass ϑ die a priori-Verteilung besitzt (z.B. aus „Erfahrung“ oder aus „Expertenwissen“).

Wir interpretieren die Likelihood-Funktion $\rho : S \times \Theta \to [0, \infty)$ als

$$\rho(x, p) = P(X = x \mid \vartheta = p)$$

(bzw. $P(X \in dx \mid \vartheta = p) = \rho(x, p)\, dx$ falls X eine Dichte besitzt)

Mit Formel von der totalen Wahrscheinlichkeit (Satz 1.41, 1.) ist

$$P(X = x) = \int_\Theta \rho(x,t)\,\alpha(t)dt$$

(bzw. $P(X \in dx) = \int_\Theta \rho(x,t)\,\alpha(t)dt\,dx$, d.h.
$P(X \le x) = \int_\Theta \int_{-\infty}^x \rho(y,t)\,dy\,\alpha(t)dt$, wenn X eine Dichte besitzt),

Mit Formel von der totalen Wahrscheinlichkeit (Satz 1.41, 1.) ist

$$P(X = x) = \int_\Theta \rho(x,t)\,\alpha(t)dt$$

(bzw. $P(X \in dx) = \int_\Theta \rho(x,t)\,\alpha(t)dt\,dx$, d.h. $P(X \leq x) = \int_\Theta \int_{-\infty}^{x} \rho(y,t)\,dy\,\alpha(t)dt$, wenn X eine Dichte besitzt),

mit der Formel von Bayes (Satz 1.41, 2.) ist

$$P(\vartheta \in dp \mid X = x) = \frac{\rho(x,p)\alpha(p)\,dp}{\int_\Theta \alpha(\vartheta')\rho(x,\vartheta')\,d\vartheta'}$$

Die *a posteriori-Dichte* (bzw. a posteriori-Gewicht, wenn ϑ diskret) bei Beobachtung x,

$$\pi_x(\vartheta) = \frac{\alpha(\vartheta)\rho(x,\vartheta)}{\int_\Theta \alpha(\vartheta')\rho(x,\vartheta')\,d\vartheta'},$$

ist die Dichte von ϑ, bedingt auf Beobachtung $X = x$, d.h.

$$P(\vartheta \leq u \mid X = x) = \int_{\Theta\cap(-\infty,u]} \pi_x(p)\,dp$$

Die *a posteriori-Dichte* (bzw. a posteriori-Gewicht, wenn ϑ diskret) bei Beobachtung x,

$$\pi_x(\vartheta) = \frac{\alpha(\vartheta)\rho(x,\vartheta)}{\int_\Theta \alpha(\vartheta')\rho(x,\vartheta')\,d\vartheta'},$$

ist die Dichte von ϑ, bedingt auf Beobachtung $X = x$, d.h.

$$P(\vartheta \le u \mid X = x) = \int_{\Theta\cap(-\infty,u]} \pi_x(p)\,dp$$

Definition 2.11

Der *Bayes-Schätzer* (zur a priori-Verteilung α) ist

$$\widehat{\vartheta}_{\mathrm{B}} = \widehat{\vartheta}_{\mathrm{B}}(x) := \mathbb{E}_{\pi_x}[\vartheta] = \int_\Theta \vartheta\,\pi_x(\vartheta)\,d\vartheta$$

(d.h. der Erwartungswert von ϑ bedingt auf $X = x$).

(Wir betrachten hier nur den Fall, dass Θ ein Intervall ist.)