

5. Elemente Statistischer Inferenz

Ziel der Statistik ist es, unter bestimmten Annahmen Aussagen über unbekannte Parameter $\theta \in \Theta$ zu machen, nachdem Beobachtungen X gemacht wurden.

Dabei unterscheidet man

- **Punktschätzungen:**

Was ist der “beste” Schätzwert $\hat{\theta}$ für den unbekannt Parameter θ ?

- **Intervallschätzungen:**

Angabe eines **Vertrauensintervalls**

Vertrauensintervalle

Zwei Arten:

- **Konfidenzintervalle** überdecken mit einer gewissen Sicherheit den unbekannt Parameter θ (bei hypothetischer Wiederholung des Zufallsexperiments).
Beachte: θ fest, X zufällig
→ **frequentistischer** Wahrscheinlichkeitsbegriff
- In einem **Kredibilitätsintervall** liegt der unbekannte Parameter mit einer gewissen Wahrscheinlichkeit.
Beachte: θ zufällig, X fest
→ **subjektivistischer** Wahrscheinlichkeitsbegriff

Beispiele

- 1 Sei $X \sim \mathcal{B}(n, \pi)$. Man beobachtet $X = 7$ bei $n = 10$ Versuchen.
 - ▶ Was ist der “beste” Schätzer $\hat{\pi}$ für den unbekannt Parameter $\theta = \pi$?
 - ▶ Wie lautet ein 95%-Vertrauensintervall?
- 2 Capture-Recapture Experiment:
Angenommen $M = 100$, $n = 50$ und $X = 33$.
 - ▶ Was ist der “beste” Schätzer \hat{N} für den unbekannt Parameter $\theta = N$?
Beachte: $N \geq N_{min} = \max(M + n - x, n)$

Beispiel 1: $\theta \in \Theta = [0, 1]$ stetig

Beispiel 2: $\theta \in \Theta = \{N_{min}, N_{min} + 1, \dots\}$ diskret

5.1 Likelihood-Inferenz

Wir haben die Wahrscheinlichkeitsfunktion $f(x)$ einer ZV X in Abhängigkeit von einem Parameter θ kennengelernt.

Beispiel:

$$X \sim B(n, \pi) \quad \Rightarrow \quad \underbrace{f(x)}_{f(x; \theta=\pi)} = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

Betrachte nun $f(x; \theta)$ als Funktion von θ für festes $X = x$:

$$\begin{aligned} L(\theta) &= f(x, \theta) && \text{heißt } \mathbf{Likelihoodfunktion} \\ l(\theta) &= \log L(\theta) && \text{heißt } \mathbf{Log-Likelihoodfunktion} \end{aligned}$$

$f(x, \theta)$ beschreibt **keine** gemeinsame Dichtefunktion.

Für iid Zufallsvariablen $x = x_1, \dots, x_n$ gilt:

$$L(\theta) = f(x, \theta) = \prod_{i=1}^n f(x_i)$$

Der Maximum-Likelihood-Schätzer

Idee: Je größer die (Log-)Likelihoodfunktion $L(\theta)$ bzw. $l(\theta)$ als Funktion von θ bei gegebenen Daten $X = x$ ist, desto “plausibler” ist der entsprechende Wert von θ .

Optimal ist somit der **Maximum-Likelihood (ML)**-Schätzer $\hat{\theta}_{ML}$, für den gelten soll:

$$\begin{aligned}L(\hat{\theta}_{ML}) &= \max_{\theta \in \Theta} L(\theta) \\ \text{bzw. } l(\hat{\theta}_{ML}) &= \max_{\theta \in \Theta} l(\theta)\end{aligned}$$

Der (ML)-Schätzer $\hat{\theta}_{ML}$ maximiert also die (Log-)Likelihoodfunktion $L(\theta)$ bzw. $l(\theta)$.

Zur Berechnung des *ML*-Schätzers

- in einfacheren Modellen analytisch möglich:
Ableitung der Log-Likelihood gleich Null setzen

Beispiel:

Binomialverteilung: $\hat{\pi}_{ML} = x/n$

- ansonsten Verwendung numerischer Algorithmen:
 - ▶ Optimierung, z.B. Funktionen `optim()` und `optimize()` in R
 - ▶ EM-Algorithmus

ML-Inferenz im Capture-Recapture-Experiment I

Im Capture-Recapture Beispiel lautet die Wahrscheinlichkeitsfunktion:

$$f(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

Dabei sind M und n bekannt. Bei beobachteter Stichprobe $X = x$ lautet die Likelihoodfunktion für den unbekannt Parameter N

$$L(\theta = N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

unter der Restriktion, dass $N \geq \max(M + n - x, n)$.

Im Unterschied zum vorhergehenden Abschnitt ist der unbekannte Parameter N nun ganzzahlig.

ML-Inferenz im Capture-Recapture-Experiment II

Man kann zeigen, dass für $x > 0$ und $\hat{N}_{naive} = \frac{Mn}{x} \notin \mathbb{N}$ gilt:

$$\hat{N}_{ML} = \text{trunc} \left\{ \frac{Mn}{x} \right\} = \text{trunc} \left\{ \hat{N}_{naive} \right\}$$

Im Fall $\hat{N}_{naive} \in \mathbb{N}$ ist der ML -Schätzer i.A. nicht eindeutig: dann wird die Likelihood sowohl durch \hat{N}_{naive} als auch durch $\hat{N}_{naive} - 1$ maximiert.

ML-Inferenz im Capture-Recapture-Experiment III

Zahlenbeispiele:

M	n	x	\hat{N}_{ML}	\hat{N}_{naive}
100	50	33	151	151.51
100	50	41	121	121.95
7	23	4	40	40.25
25	30	10	{74, 75}	75
13	10	5	{25, 26}	26

In den beiden Extremfällen erhält man:

M	n	x	\hat{N}_{ML}	\hat{N}_{naive}
100	50	0	∞	∞
100	50	50	100	100

Invarianz des ML -Schätzers

Wichtige und nützliche Eigenschaft des ML -Schätzers!

Sei $\hat{\theta}_{ML}$ der ML -Schätzer für θ und $\varphi = \varphi(\theta)$ eine beliebige (eindeutige) Funktion von θ . Dann ist der ML -Schätzer von φ :

$$\hat{\varphi}_{ML} = \varphi(\hat{\theta}_{ML})$$

Beispiel: Bestimmung des ML -Schätzers für die Chance $\gamma = \frac{\pi}{1-\pi}$ im Binomialexperiment:

$$\hat{\gamma}_{ML} = \frac{\hat{\pi}_{ML}}{1 - \hat{\pi}_{ML}} = \frac{\frac{x}{n}}{1 - \frac{x}{n}} = \frac{x}{n - x}$$

Das Testproblem

Wir sind nun daran interessiert, Aussagen darüber zu treffen, in welchen Teilmengen des Parameterraumes Θ sich der feste, aber unbekannte, Parameter $\theta \in \Theta$ mutmaßlich befindet.

Dazu unterteilen wir den Parameterraum Θ in zwei disjunkte Teilmengen Θ_0 und Θ_1 mit $\Theta = \Theta_0 \cup \Theta_1$, wobei $\Theta_0 \cap \Theta_1 = \emptyset$. Es ist nun eine Entscheidungsregel gesucht, für welche der beiden Zustände $\theta \in \Theta_0$ oder $\theta \in \Theta_1$ wir uns basierend auf einem Experiment (also Daten $\mathbf{X} = (X_1, \dots, X_n)$) entscheiden sollen.

Θ_0 heißt Hypothese oder Null-Hypothese. Θ_1 heißt Alternative. Die Formulierung

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1$$

heißt Testproblem.

Der statistische Test

Eine Entscheidungsregel, um zwischen H_0 bzw. H_1 zu entscheiden, nennt man einen statistischen Test.

Eine Funktion $\psi : \mathbb{R}^n \rightarrow \{0, 1\}$ heißt Test für H_0 gegen H_1 . Wenn für den Erwartungswert der Funktion ψ gilt: $E(\psi) = P(\psi = 1) \leq \alpha, \alpha \in [0, 1]$, für alle $\theta \in \Theta_0$ (also in der Null-Hypothese), so heißt ψ Niveau- α -Test.

Fehlerarten:

- H_0 richtig und $\psi = 0$: OK!
- H_1 richtig und $\psi = 1$: OK!
- H_0 richtig und $\psi = 1$: Fehler 1. Art!
- H_1 richtig und $\psi = 0$: Fehler 2. Art!

Test und Teststatistik

In aller Regel ist ein Test ψ von der Form

$$\psi(\mathbf{X}) = I(T(\mathbf{X}) \geq c)$$

Die Funktion $T : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt Teststatistik und die Konstante c heißt kritischer Wert (der überschritten werden muss, um sich gegen die Null-Hypothese zu entscheiden).

Der kritische Wert kann so bestimmt werden, dass ψ ein Niveau- α -Test ist. Es gilt:

$$E_0(\psi(\mathbf{X}) = 1) = E_0(I(T(\mathbf{X}) \geq c)) = P_0(T(\mathbf{X}) \geq c) \leq \alpha.$$

Also ist c das $1 - \alpha$ -Quantil der Verteilung von $T(\mathbf{X})$ wenn $\theta \in \Theta_0$ und wird deshalb auch mit $c_{1-\alpha}$ bezeichnet.

Test und Teststatistik

Ein Test ψ ist von der Form

$$\psi(\mathbf{X}) = I(T(\mathbf{X}) \in \mathcal{C})$$

Die Funktion $T : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt Teststatistik und die Menge \mathcal{C} heißt kritische Region (Ablehnungsbereich) des Tests.

Der kritische Bereich kann so bestimmt werden, dass ψ ein Niveau- α -Test ist. Es gilt:

$$E_0(\psi(\mathbf{X}) = 1) = P_0(\psi(\mathbf{X}) = 1) = P_0(T(\mathbf{X}) \in \mathcal{C}) \leq \alpha.$$

Ablehnungsbereich

In der Regel lautet der Test ψ :

$\psi(\mathbf{X}) = I(T(\mathbf{X}) > c_{1-\alpha})$ falls große Werte von $T(\mathbf{X})$ gegen H_0 sprechen.

$\psi(\mathbf{X}) = I(T(\mathbf{X}) < c_\alpha)$ falls kleine Werte von $T(\mathbf{X})$ gegen H_0 sprechen.

$\psi(\mathbf{X}) = I(T(\mathbf{X}) < c_{\alpha/2} \text{ oder } T(\mathbf{X}) > c_{1-\alpha/2})$
falls große und kleine Werte von $T(\mathbf{X})$ gegen H_0 sprechen.

Die Schranken sind die Quantile der Verteilung von $T(\mathbf{X})$ wenn $\theta \in \Theta_0$.

$c_{1-\alpha}$: $1 - \alpha$ -Quantil.

c_α : α -Quantil

$c_{\alpha/2}$: $\alpha/2$ -Quantil

$c_{1-\alpha/2}$: $1 - \alpha/2$ -Quantil

Gütefunktion

Die Wahrscheinlichkeit für den Fehler 2. Art (H_1 richtig aber $\psi = 0$),

$$P(\psi = 0 | \theta \in \Theta_1) = \beta$$

ist vom wahren Parameter θ (bzw. von der Differenz $\theta - \hat{\theta}$) abhängig.

Die Gütefunktion

$$g(\theta) = P(\psi = 1 | \theta)$$

gibt die Wahrscheinlichkeit H_0 zu verwerfen in Abhängigkeit von θ an. Sie besitzt die folgenden Eigenschaften:

- Für $\theta \in \Theta_0$ gilt $g(\theta) \leq \alpha$.
- Für $\theta \in \Theta_1$ ist $1 - g(\theta) = \beta$.
- Für $\theta \in \Theta_1$ wird $g(\theta)$ auch Power oder Trennschärfe eines Tests bezeichnet.

Beispiel: Faire Münze

$$H_0 : \pi = \frac{1}{2} \text{ vs. } H_1 : \pi \neq \frac{1}{2}$$

Teststatistik $T(x) = x$; lehne die Null-Hypothese ab, wenn x zu groß *oder* zu klein ist (zweiseitiger Test).

Der Ausgang unseres Experimentes sei $x = 7$ bei $n = 10$. Wir bestimmen jetzt das $\alpha/2$ -Quantil und das $1 - \alpha/2$ -Quantil der Verteilung von $T(x)$ unter H_0 , also der Verteilung $\mathcal{B}(n, 0.5)$ für $\alpha = 0.05$:

```
> qbinom(0.05/2, size = 10, prob = 0.5)      # 2.5%-Quantil
```

```
[1] 2
```

```
> qbinom(1 - 0.05/2, size = 10, prob = 0.5) # 97.5%-Quantil
```

```
[1] 8
```

Damit können wir H_0 nicht ablehnen, da $x = 7$ weder zu klein (kleiner 2) noch zu groß (größer 8) ist.

Likelihood-Quotienten-Tests

Eine formale Möglichkeit, Teststatistiken zu konstruieren, ist die Anwendung des Likelihood-Quotienten-Prinzips: Der Quotient

$$\frac{\max_{\Theta_0} L(\theta)}{\max_{\Theta} L(\theta)} = \frac{\max_{\Theta_0} L(\theta)}{L(\hat{\theta}_{ML})}$$

heißt Likelihood-Quotient und steht in engem Zusammenhang zur normierten (Log)-Likelihoodfunktion.

Die normierte Likelihoodfunktion

Den Wert der (Log-) Likelihoodfunktion $L(\theta)$ bzw. $l(\theta)$ am ML -Schätzer kann man nicht interpretieren. Daher verwendet man gerne die **normierte (Log-) Likelihoodfunktion**:

$$\tilde{L}(\theta) = \frac{L(\theta)}{L(\hat{\theta}_{ML})}$$

$$\tilde{l}(\theta) = l(\theta) - l(\hat{\theta}_{ML})$$

Es gilt: $0 \leq \tilde{L}(\theta) \leq 1$ und $-\infty \leq \tilde{l}(\theta) \leq 0$

Likelihood-Intervalle

Zur Bestimmung von Konfidenzintervallen muss man nun entscheiden, welche Werte von $\tilde{l}(\theta)$ zu “unplausibel” sind.

Dazu kann man den Test ‘invertieren’: nicht mehr zu einem festen Θ_0 einen Niveau- α -Test durchführen, sondern suche Menge Θ_0 , für die der Niveau- α -Test ψ gerade *nicht* ablehnt, also:

$$\{\theta \in \Theta : \psi(\mathbf{X}) = 0\} = \{\theta \in \Theta : T(\mathbf{X}) < c_{1-\alpha}\}$$

Likelihood-Intervalle II

Man kann unter bestimmten Annahmen zeigen, dass für einen bestimmten Schwellenwert $c = c(\alpha)$

$$\{\theta : \tilde{l}(\theta) \geq c\}$$

bzw. $\{\theta : \tilde{L}(\theta) \geq \exp(c)\}$

ein **Konfidenzintervall** für θ (**Likelihood-Intervall**) zum approximativen **Niveau** $1 - \alpha$ ist.

Interpretation: Bei hypothetischer Wiederholung des zugrundeliegenden Zufallsexperiments überdecken die so konstruierten Likelihood-Intervalle in ungefähr $(1 - \alpha) \cdot 100\%$ aller Fälle den unbekannt Parameter θ .

Likelihood-Intervalle III

Die Werte von c sind folgender Tabelle zu entnehmen:

$1 - \alpha$	c	$\exp(c)$
0.9	-1.33	0.259
0.95	-1.92	0.147
0.99	-3.32	0.036

Die Bestimmung von Likelihood-Intervallen ist nur numerisch möglich, dort aber einfach durchzuführen.

Likelihood-Intervalle sind (wie der *ML*-Schätzer) invariant bzgl. monotonen Transformationen des Parameters.

Beispiel: Binomialverteilung

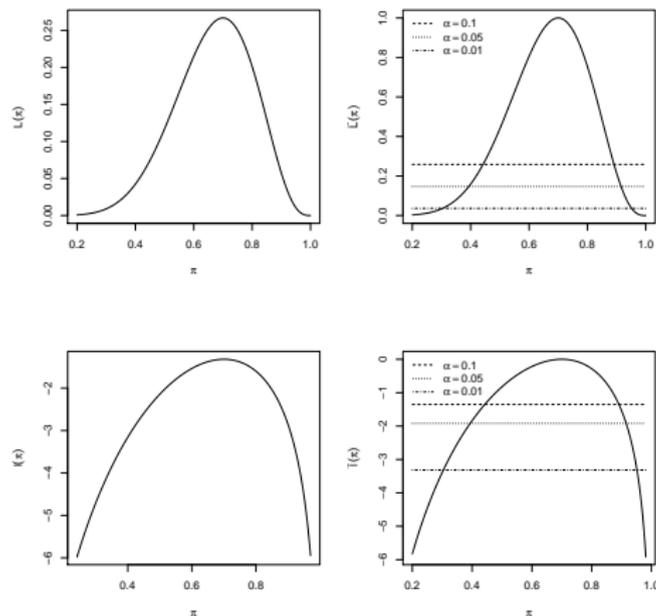
Sei $X \sim \mathcal{B}(n, \pi)$ mit $n = 10$ und $x = 7$.

Damit erhält man als *ML*-Schätzer: $\hat{\pi}_{ML} = 0.7$

$1 - \alpha$	Likelihood-Intervall für π
0.9	[0.44; 0.89]
0.95	[0.39; 0.92]
0.99	[0.30; 0.95]

Beachte: Die Konfidenzintervalle sind i.A. nicht symmetrisch um $\hat{\pi}_{ML}$.

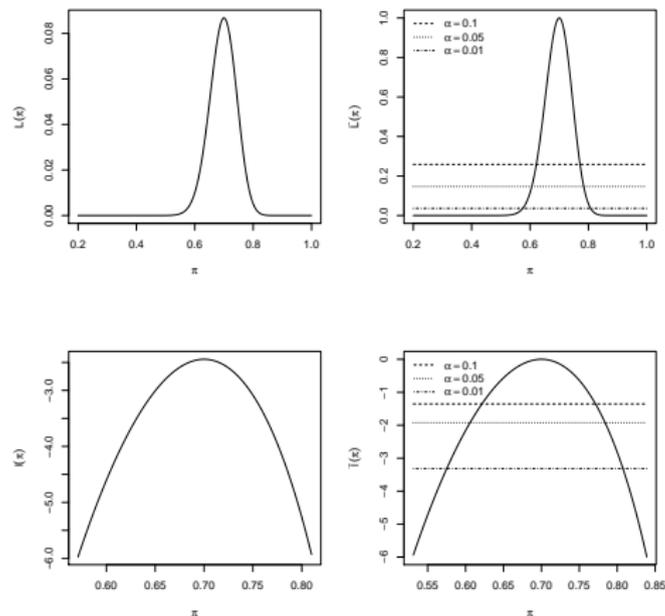
Beispiel: Binomialverteilung ($n = 10, x = 7$)



Linien verdeutlichen
Likelihood-Intervalle zu
unterschiedlichen Niveaus.

Likelihood (oben links), normierte Likelihood
(oben rechts), Loglikelihood (unten links) und
normierte Loglikelihood (unten rechts)

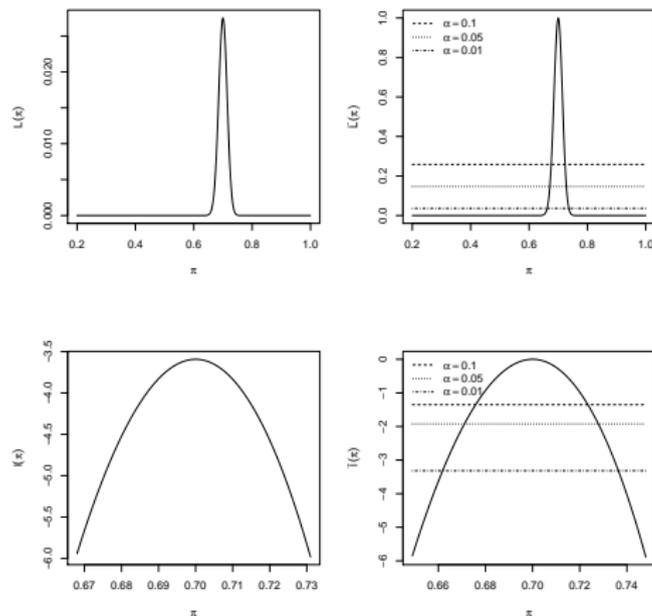
Beispiel: Binomialverteilung ($n = 100, x = 70$)



Linien verdeutlichen
Likelihood-Intervalle zu
unterschiedlichen Niveaus.

Likelihood (oben links), normierte Likelihood
(oben rechts), Loglikelihood (unten links) und
normierte Loglikelihood (unten rechts)

Beispiel: Binomialverteilung ($n = 1000$, $x = 700$)



Linien verdeutlichen
Likelihood-Intervalle zu
unterschiedlichen Niveaus.

Likelihood (oben links), normierte Likelihood
(oben rechts), Loglikelihood (unten links) und
normierte Loglikelihood (unten rechts)

Quadratische Approximation der Log-Likelihood

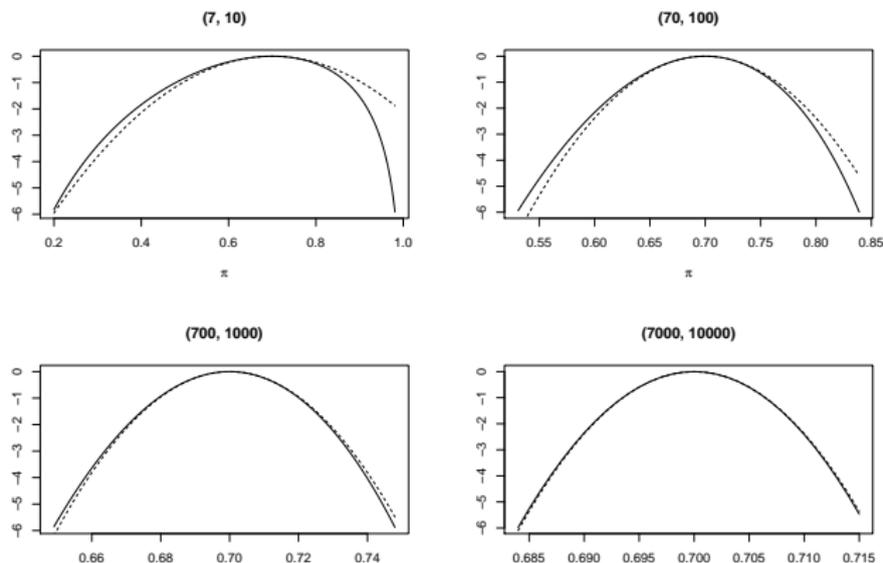
Man kann mit Hilfe einer Taylorreihendarstellung um $\hat{\theta}_{ML}$ zeigen, dass die normierte Loglikelihoodfunktion $\tilde{l}(\theta)$ approximativ eine quadratische Funktion ist:

$$\tilde{l}(\theta) \approx \frac{1}{2} \cdot l''(\hat{\theta}_{ML}) \cdot (\theta - \hat{\theta}_{ML})^2$$

Hier ist $l''(\hat{\theta}_{ML})$ ($= \tilde{l}''(\hat{\theta}_{ML})!$) die zweite Ableitung (die Krümmung) von $l(\theta)$, ausgewertet am ML -Schätzer.

Beachte: Die quadratische Approximation wird umso besser, je mehr Daten der Likelihood zugrundeliegen.

Beispiel: Binomialverteilung



Vergleich der normierten Loglikelihood mit der quadratischen Approximation für $X \sim \mathcal{B}(n, \pi)$:

$n = 10, X = 7$ (oben links), $n = 100, X = 70$ (oben rechts),
 $n = 1000, X = 700$ (unten links) und $n = 10000, X = 7000$ (unten rechts)

Der Standardfehler

Durch Einsetzen der quadratischen Approximation für $\tilde{l}(\theta)$ in

$$\{\theta : \tilde{l}(\theta) \geq c\}$$

erhält man

$$\hat{\theta}_{ML} \pm \sqrt{-2c} \cdot \sqrt{\left[-l''(\hat{\theta}_{ML})\right]^{-1}}$$

als Konfidenzintervall (**Wald-Intervall**) zum approximativen Niveau $1 - \alpha$.
Daher definiert man den **Standardfehler** ("Standard Error") als

$$SE(\hat{\theta}_{ML}) := \sqrt{\left[-l''(\hat{\theta}_{ML})\right]^{-1}}$$

Der Standardfehler des ML-Schätzers

$$SE(\hat{\theta}_{ML}) := \sqrt{\left[-I''(\hat{\theta}_{ML})\right]^{-1}}$$

kann als Schätzung der Standardabweichung des ML-Schätzers (im frequentistischen Sinne) angesehen werden.

Ebenso ist $\left[-I''(\hat{\theta}_{ML})\right]^{-1}$ eine Schätzung der Varianz des ML-Schätzers.

Asymptotische Eigenschaften des ML -Schätzers

Man kann zeigen, dass (unter Regularitätsbedingungen) asymptotisch (für großen Stichprobenumfang) gilt:

$$\hat{\theta}_{ML} \overset{a}{\sim} \mathcal{N}(\mu = \theta, \sigma^2 = SE(\hat{\theta}_{ML})^2)$$

Nach Standardisierung erhält man:

$$\tilde{\theta}_{ML} = \frac{\hat{\theta}_{ML} - \theta}{SE(\hat{\theta}_{ML})} \overset{a}{\sim} \mathcal{N}(0, 1)$$

D.h., der ML -Schätzer ist asymptotisch **unverzerrt** und **normalverteilt** mit Standardabweichung gleich dem Standardfehler.

Motivation von Wald-Intervallen

Sei $x_\alpha = \Phi^{-1}(\alpha)$ das α -Quantil der Standardnormalverteilung. Dann gilt mit einer Wahrscheinlichkeit von $\beta = 1 - \alpha$, dass $\tilde{\theta}_{ML}$ asymptotisch im Intervall $[x_{\alpha/2}, x_{1-\alpha/2}]$ ist.

Beachte: Wegen der Symmetrie der Normalverteilung ist $x_{\alpha/2} = -x_{1-\alpha/2}$

Wald-Intervalle

Das Wald-Intervall zum Niveau $1 - \alpha$ ist also:

$$\hat{\theta}_{ML} \pm d \cdot SE(\hat{\theta}_{ML})$$

Die Werte von d sind folgender Tabelle zu entnehmen:

$1 - \alpha$	c	$d = \sqrt{-2c}$
0.9	-1.35	1.65
0.95	-1.92	1.96
0.99	-3.32	2.58

$$d = x_{1-\alpha/2}$$

Beispiel: Binomialverteilung

Es ergibt sich

$$SE(\hat{\pi}_{ML}) = \sqrt{\frac{\hat{\pi}_{ML}(1 - \hat{\pi}_{ML})}{n}}$$

Beispiel: Für $X = 7$ und $n = 10$ ergibt sich: $SE(\hat{\pi}_{ML}) = 0.145$
→ Tabelle mit Wald-Intervallen:

$1 - \alpha$	Wald-Intervall
0.9	[0.46; 0.94]
0.95	[0.42; 0.98]
0.99	[0.33; 1.07]

Eigenschaften von Wald-Intervallen

- Wald-Intervalle sind immer symmetrisch um den ML -Schätzer; Invarianzeigenschaft geht verloren
- Wald-Intervalle sind einfacher zu berechnen als Likelihood-Intervalle, haben aber (leicht) schlechtere theoretische Eigenschaften
- im Beispiel:
offensichtliches Problem für $1 - \alpha = 0.99$: obere Grenze ist größer als 1!
- für n groß werden Wald-Intervalle Likelihood-Intervallen immer ähnlicher

Likelihood-Inferenz für stetige ZVn

Likelihood-Inferenz lässt sich analog zu diskreten Zufallsvariablen auch bei stetigen Zufallsvariablen anwenden.

Beispiel:

X_1, X_2, \dots, X_n seien unabhängige Beobachtungen aus einer $\mathcal{E}(\lambda)$ -Verteilung. Wie lautet der *ML*-Schätzer von $\theta = \lambda$ und dessen Standardfehler? Für $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$ ergibt sich (Herleitung in Vorlesung):

$$\begin{aligned}\hat{\lambda}_{ML} &= 1/\bar{x} \\ SE(\hat{\lambda}_{ML}) &= 1/(\sqrt{n}\bar{x}) = \hat{\lambda}_{ML}/\sqrt{n}\end{aligned}$$

→ 95% Wald-Intervall für λ : $1/\bar{x} \pm 1.96 \cdot 1/(\sqrt{n}\bar{x})$

Wie gut ist die Approximation?

Dazu wird eine Simulationsstudie mit wahrem Wert $\theta = \lambda$ und variierendem n durchgeführt:

- Berechnung von m Konfidenzintervallen, jeweils basierend auf n exponentialverteilten Zufallsvariablen
- Berechnung der empirischen **Verteilung** des ML -Schätzers und der **Überdeckungshäufigkeit** (Anteil der Konfidenzintervalle, die den wahren Wert beinhalten)

Mehr zur Asymptotik des ML -Schätzers

Betrachte wieder exemplarisch die unabhängigen X_1, X_2, \dots, X_n aus einer $\mathcal{E}(\lambda)$ -Verteilung. Es gilt:

$$\begin{aligned}\mu = E(X_i) &= 1/\lambda \\ \text{Var}(X_i) &= 1/\lambda^2\end{aligned}$$

Wir wollen nun $\mu = 1/\lambda$ durch Maximum Likelihood schätzen. Es ergibt sich:

$$\begin{aligned}\hat{\mu}_{ML} &= \bar{x} \quad (\text{wegen Invarianzeigenschaft}) \\ SE(\hat{\mu}_{ML}) &= \bar{x}/\sqrt{n} = \hat{\mu}_{ML}/\sqrt{n} \quad (\text{ohne Beweis})\end{aligned}$$

→ 95% Wald-Intervall für μ : $\bar{x} \pm 1.96 \cdot \bar{x}/\sqrt{n}$

Der ML -Schätzer als Zufallsvariable

Betrachte $\bar{X} = \frac{1}{n} \sum X_i$ mit unabhängigen $X_i \sim \mathcal{E}(\lambda)$. Es folgt:

$$\begin{aligned} E(\bar{X}) &= \mu \\ \text{Var}(\bar{X}) &= \mu^2/n = (\mu/\sqrt{n})^2 \end{aligned}$$

d.h. der ML -Schätzer \bar{X} ist erwartungstreu mit Standardabweichung μ/\sqrt{n} .
Daher gilt wegen dem zentralen Grenzwertsatz (ZGWS):

$$\bar{X} \stackrel{a}{\sim} \mathcal{N}(\mu, \sigma^2 = \mu^2/n)$$

Konfidenzintervall basierend auf dem ZGWS

Ein 95%-Konfidenzintervall für μ basierend auf dem ZGWS ist daher

$$\bar{x} \pm 1.96 \cdot \mu / \sqrt{n}$$

Problem:

μ ist unbekannt und deshalb verwendet man eine “*plug-in*”-Schätzung von μ durch $\hat{\mu}_{ML} = \bar{x}$. Damit erhält man:

$$\bar{x} \pm 1.96 \cdot \bar{x} / \sqrt{n}$$

Dieses Intervall ist identisch zu dem Wald-Intervall, welches auf dem Standardfehler basiert!

Also: Der **Standardfehler** ist ein empirischer Schätzer der **Standardabweichung** des *ML*-Schätzers.

Bemerkungen

- exakte Analogie funktioniert nicht immer:

z.B. für $\hat{\lambda}_{ML} = 1/\bar{x}$ mit $SE(\hat{\lambda}_{ML}) = \hat{\lambda}_{ML}/\sqrt{n}$ gilt (ohne Beweis):

$$E(1/\bar{X}) = \lambda \cdot \frac{n}{n-1}$$

$$\text{Var}(1/\bar{X}) = \frac{\lambda^2}{(n-2)} \cdot \frac{n^2}{(n-1)^2}$$

- die Analogie gilt aber zumindest asymptotisch!
- in folgenden Beispielen gilt die Analogie exakt:
 - ▶ Binomialesperiment: $\hat{\pi}_{ML} = \bar{x}$
 - ▶ *ML*-Schätzung des Parameters q im Hardy-Weinberg-Gleichgewicht (wird in den nächsten Folien skizziert)

ML-Inferenz im H-W-Gleichgewicht

Angenommen wir beobachten $\mathbf{x} = (x_1, x_2, x_3) = (600, 320, 80)$ bei $n = 1000$.

Ziel:

ML-Schätzung des Parameters $q = \pi_1 + \pi_2/2$ unter Annahme einer Trinomialverteilung mit Wahrscheinlichkeiten $\pi_1 = q^2$, $\pi_2 = 2q(1 - q)$ und $\pi_3 = (1 - q)^2$ (H-W-Gleichgewicht)

Wir wissen:

Die ML-Schätzer von π_1 und π_2 sind $\hat{\pi}_1 = x_1/n$ und $\hat{\pi}_2 = x_2/n$. Wegen der Invarianzeigenschaft gilt:

$$\hat{q}_{ML} = \frac{x_1 + x_2/2}{n}$$

ML-Inferenz im H-W-Gleichgewicht II

Man kann zeigen, dass die zugehörige Zufallsvariable $\frac{X_1+X_2}{n}$ die Varianz $\frac{1}{2} q(1-q)$ hat.

Andererseits kann gezeigt werden, dass der Standardfehler von \hat{q}_{ML} folgenden Wert hat:

$$SE(\hat{q}_{ML}) = \sqrt{\frac{1}{2} \hat{q}_{ML} (1 - \hat{q}_{ML})}$$

Auch hier erhält man den Standardfehler durch “plug-in” des *ML*-Schätzers in die Formel für die Varianz des *ML*-Schätzers.

ML-Inferenz im H-W-Gleichgewicht III

Für die Daten $\mathbf{x} = (x_1, x_2, x_3) = (600, 320, 80)$ und $n = 1000$ gilt also:

$$\hat{q}_{ML} = \frac{600 + 320/2}{1000} = 0.76$$

Daraus lassen sich die erwarteten Anzahlen berechnen:

$$E = n \cdot (\hat{q}_{ML}^2, 2\hat{q}_{ML}(1 - \hat{q}_{ML}), (1 - \hat{q}_{ML})^2) = (577.6, 364.8, 57.6)$$

Frage: Ist der Unterschied zwischen erwarteten und beobachteten Anzahlen “zufällig” oder deutet er darauf hin, dass die Population nicht im H-W-Gleichgewicht ist? ...

später: Modellanpassung, Goodness-of-Fit.

Wiederholung: Quadrat einer Standardnormalverteilung

Wie lautet die Dichte von $Y = X^2$, falls $X \sim \mathcal{N}(0, 1)$?

Es wurde berechnet, dass

$$f(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} y\right)$$

Dies entspricht der Dichte einer $\mathcal{G}(.5, .5)$, also einer χ^2 -Verteilung mit 1 Freiheitsgrad: $Y = X^2 \sim \chi_1^2$

Motivation von Likelihood-Intervallen

Betrachte die Taylor-Approximation der Log-Likelihood:

$$l(\theta) \approx l(\hat{\theta}_{ML}) - \frac{1}{2} \frac{(\hat{\theta}_{ML} - \theta)^2}{SE(\hat{\theta}_{ML})^2} = l(\hat{\theta}_{ML}) - \frac{1}{2} \tilde{\theta}_{ML}^2$$

Wegen $\tilde{\theta}_{ML} \stackrel{a}{\sim} \mathcal{N}(0, 1)$ folgt:

$$2 \log \frac{L(\hat{\theta}_{ML})}{L(\theta)} = -2\tilde{l}(\theta) \stackrel{a}{\sim} \chi_1^2$$

Motivation von Likelihood-Intervallen II

Die kritischen Werte c in den Formeln zur Berechnung von Likelihood-Intervallen

$$\{\theta : \tilde{l}(\theta) \geq c\}$$

bzw. $\{\theta : \tilde{L}(\theta) \geq \exp(c)\}$

ergeben sich einfach durch Transformation der entsprechenden Quantile $x_\alpha = F^{-1}(\alpha)$ der χ_1^2 -Verteilung:

$$c = -\frac{x_{1-\alpha}}{2}$$

5.2 Erwartungstreue

Eine wünschenswerte Eigenschaft von Punktschätzern (im frequentistischen Sinne) ist die **Erwartungstreue**:

Ein Schätzer $\hat{\theta}$ (als Funktion der zufälligen Stichprobe X) heißt **erwartungstreu** oder **unverzerrt** für einen unbekanntem Parameter θ , falls gilt:

$$E(\hat{\theta}) = \theta$$

Beispiel: Die relative Häufigkeit $\hat{\pi} = X/n$ ist ein erwartungstreuer Schätzer der Wahrscheinlichkeit π im Binomialverteilungsmodell.

Bemerkungen zur Erwartungstreue

- Die Erwartungstreue ist nicht invariant bzgl. monotonen Transformationen! Das heißt, ist $\hat{\theta}$ erwartungstreu für θ , so ist $g(\hat{\theta})$ im Allgemeinen nicht erwartungstreu für $g(\theta)$.
- Die Existenz von erwartungstreuen Schätzern ist nicht gesichert.
- Erwartungstreue Schätzer sind nicht notwendigerweise Element des Parameterraums Θ .
- *ML*-Schätzer sind nicht immer erwartungstreu, zumindest aber immer *asymptotisch* erwartungstreu, d.h. für wachsenden Stichprobenumfang im Grenzwert erwartungstreu.

Beispiel: Taxis in Lübeck I

Alle Taxis in Lübeck seien von $1, \dots, N$ durchnummeriert. Ein Besucher sieht an einem Taxistand $n = 3$ Taxis und fragt nach deren Nummern:

$$Y = \{Y_1, \dots, Y_n\}$$

Wie kann er daraus einen erwartungstreuen Schätzer für $\theta = N$ berechnen?
Betrachte $X = \max(Y)$.

Man kann zeigen (Vorlesung), dass

$$\hat{N} = \frac{n+1}{n} \max(Y) - 1$$

ein erwartungstreuer Schätzer für N ist.

Beispiel: Taxis in Lübeck II

Die Likelihoodfunktion ist

$$L(N) = \text{const} \cdot \frac{(N - n)!}{N!} \text{ für } N = x, x + 1, \dots$$

Diese wird maximiert für $N = x$! Das heißt der *ML*-Schätzer für N ist $\max(Y)$, der kleinstmögliche Wert!

Bias & Mean Squared Error (MSE)

Der Bias eines Schätzers $\hat{\theta}$ ist wie folgt definiert:

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Er beschreibt also die systematische Abweichung des Schätzers.

$\text{Bias}(\hat{\theta}) = 0 \Leftrightarrow \hat{\theta}$ ist erwartungstreu.

In vielen Situationen kommt es zu einem "Bias-Varianz-Trade-Off".
Darum wird häufig der MSE zum Vergleich von mehreren Schätzfunktionen herangezogen.

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2$$

5.3 Bayes-Inferenz

- alternativer Ansatz zur statistischen Inferenz
- basiert auf subjektivistischem Wahrscheinlichkeitskonzept
- unbekannter Parameter θ ist nun eine Zufallsvariable, versehen mit einer Wahrscheinlichkeitsfunktion $f(\theta)$
- wir behandeln hier zunächst nur diskrete Parameter

Satz von Bayes für Wahrscheinlichkeitsfunktionen

Seien X und Y zwei Zufallsvariablen mit gemeinsamer Wahrscheinlichkeitsfunktion $f_{X,Y}(x,y)$ und daraus abgeleiteten Wahrscheinlichkeitsfunktionen $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$ und $f_{Y|X}(y|x)$. Dann gilt für alle x und alle y mit $f(y) > 0$:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\sum_x f_{Y|X}(y|x)f_X(x)}$$

Dies folgt direkt aus der Definition der bedingten Wahrscheinlichkeitsfunktion.

Bayes-Inferenz

Sei $X = x$ eine Beobachtung eines Zufallsexperiments, das von einem unbekanntem Parameter $\theta \in \Theta$ abhängt, wobei Θ abzählbar sei.

Dann gilt mit dem Satz von Bayes für Wahrscheinlichkeitsfunktionen:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\sum_{\theta} f(x|\theta)f(\theta)}$$

Beachte: Da $X = x$ beobachtet wurde, muss automatisch $P(X = x) > 0$ gelten.

Beispiel: Posteriori-Verteilung im Binomialexperiment

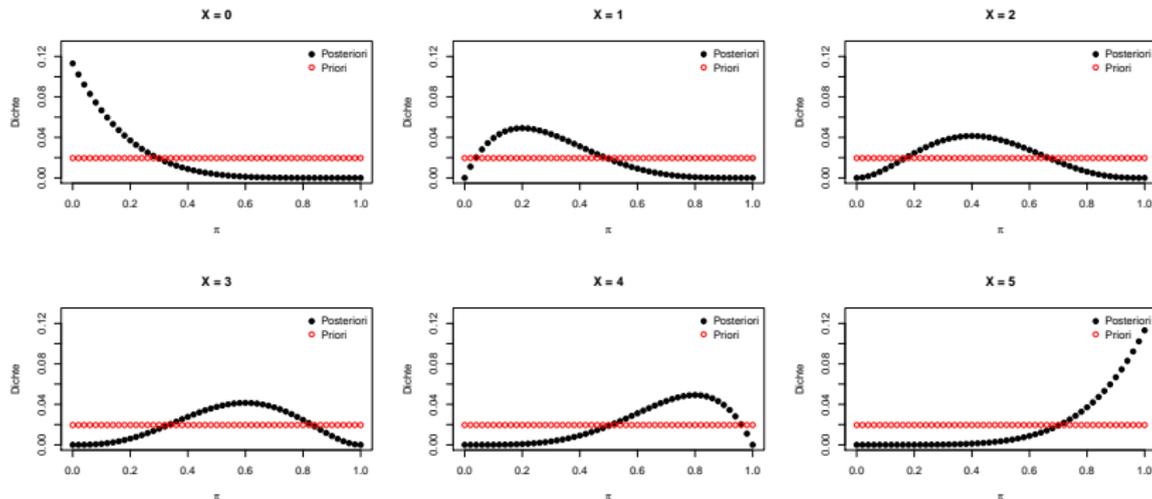
Angenommen wir interessieren uns für den Parameter $\theta = \pi$, nehmen aber an, dass nur Werte in $\Pi = \{0.00, 0.02, 0.04, \dots, 0.98, 1.00\}$ erlaubt sind.

Als **Priori-Verteilung** $f(\pi)$ können wir z.B. eine Gleichverteilung auf den 51 Elementen von Π wählen, also $f(\pi) = 1/51$ für alle $\pi \in \Pi$.

Nach der Beobachtung $X = x$ aus einer $\mathcal{B}(n, \pi)$ -Verteilung ergibt sich die **Posteriori-Verteilung**

$$f(\pi|x) = \frac{f(x|\pi)f(\pi)}{\sum_{\pi} f(x|\pi)f(\pi)}$$

Beispiel: Binomialverteilung ($n = 5$)



Posteriori-Verteilung im Binomialexperiment für $X \sim \mathcal{B}(5, \pi)$ bei Priori-Gleichverteilung.

Eine Dreiecksverteilung als Priori-Verteilung

Idee: favorisiere Werte von π nahe bei 0.5

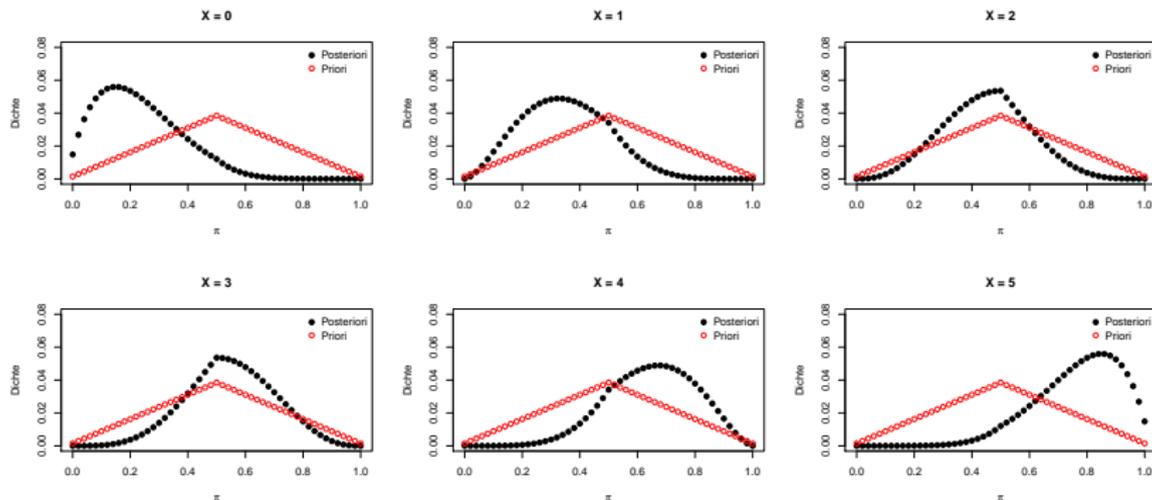
Wähle z.B.

$$f(\pi) = \frac{1}{C} \{26 - 25 \cdot |2 \cdot \pi - 1|\}$$

für $\pi \in \{0.00, 0.02, 0.04, \dots, 0.98, 1.00\}$ und $C = \frac{1}{676}$.

Bemerkung: C ist so gewählt, dass $\sum_{\pi} f(\pi) = 1$ gilt.

Beispiel: Binomialverteilung ($n = 5$)



Posteriori-Verteilung im Binomialexperiment für $X \sim \mathcal{B}(5, \pi)$ bei
Priori-Dreiecksverteilung.

Bayesianische Punktschätzer

Als Punktschätzer bietet sich nun ein Lageparameter der Posteriori-Verteilung an, z.B. der

- Posteriori-Erwartungswert $\hat{\theta}_{\text{Erw}} = E(\theta|x) = \sum_{\theta \in \Theta} \theta f(\theta|x)$
- Posteriori-Modus $\hat{\theta}_{\text{Mod}} = \arg \max_{\theta \in \Theta} f(\theta|x)$
- Posteriori-Median $\hat{\theta}_{\text{Med}} = \min\{\theta \in \Theta : F(\theta|x) \geq 0.5\}$

wobei $F(\theta|x)$ die Verteilungsfunktion der Posteriori-Verteilung mit Wahrscheinlichkeitsfunktion $f(\theta|x)$ ist.

Posteriori-Modus bei Priori-Gleichverteilung

Bei Priori-Gleichverteilung ist der Posteriori-Modus $\hat{\theta}_{\text{Mod}}$ gleich dem *ML*-Schätzer $\hat{\theta}_{\text{ML}}$, da

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\sum_{\theta} f(x|\theta)f(\theta)} = \frac{f(x|\theta)}{\sum_{\theta} f(x|\theta)}$$

Da der Nenner $\sum_{\theta} f(x|\theta)$ nicht von θ abhängt, folgt:

$$\hat{\theta}_{\text{Mod}} = \arg \max_{\theta \in \Theta} f(\theta|x) = \arg \max_{\theta \in \Theta} f(x|\theta) = \hat{\theta}_{\text{ML}}$$

Bayesianische Punktschätzer im Beispiel

Bei Priori-Gleichverteilung (G) bzw. Dreiecksverteilung (D) ergeben sich folgende Punktschätzer bei Beobachtung von $X = x$ Erfolgen bei $n = 5$ Versuchen:

x	Erwartungswert		Modus		Median	
	G	D	G	D	G	D
0.00	0.13	0.23	0.00	0.16	0.10	0.22
1.00	0.29	0.35	0.20	0.32	0.26	0.34
2.00	0.43	0.45	0.40	0.50	0.42	0.46
3.00	0.57	0.55	0.60	0.50	0.58	0.54
4.00	0.71	0.65	0.80	0.68	0.74	0.66
5.00	0.87	0.77	1.00	0.84	0.90	0.78

Bayesianische Intervallschätzer

- Prinzipiell ist jede Teilmenge A des Trägers Θ , für die gilt

$$\sum_{\theta \in A} f(\theta|x) \geq 1 - \alpha$$

eine Kredibilitätsregion (bzw. -intervall) zum Niveau $1 - \alpha$.

- Zusätzlich kann man noch fordern, dass für alle $\theta_1 \in A$ und $\theta_2 \in \Theta \setminus A$ gelten muss:

$$f(\theta_1|x) \geq f(\theta_2|x)$$

⇒ “highest posterior density region” (**HPD-Region**)

Berechnung von HPD-Regionen

- 1 Sortiere die Werte der Posteriori-Verteilung $f(\theta|x)$ der Größe nach (absteigend).
- 2 Summiere die Werte kumulativ auf (Funktion `cumsum()` in R), bis die Summe größer als das Niveau $1 - \alpha$ ist.

⇒ Die entsprechenden Werte von θ definieren dann eine HPD-Region.

HPD-Regionen im Beispiel

Bei Priori-Gleichverteilung (G) bzw. Dreiecksverteilung (D) ergeben sich folgende 95% HPD-Regionen bei Beobachtung von $X = x$ Erfolgen bei $n = 5$ Versuchen:

x	95% HPD-Region für π	
	G	D
0	[0.00;0.36]	[0.00;0.46]
1	[0.02;0.56]	[0.08;0.60]
2	[0.12;0.74]	[0.18;0.72]
3	[0.26;0.88]	[0.28;0.82]
4	[0.44;0.98]	[0.40;0.92]
5	[0.64;1.00]	[0.54;1.00]

Beispiel: Capture-Recapture-Experiment

Betrachte den unbekannt Parameter $\theta = N$ als diskrete ZV mit Träger

$$\mathcal{T} = \{M, M + 1, \dots, Y_{\max}\}$$

Beachte: **Vor** der Beobachtung $X = x$ (bei bekanntem $n!$) weiss man nur, dass mindestens M Fische im See schwimmen.

Lege nun eine Priori-Verteilung fest, z.B. eine Gleichverteilung

$$f(N) = \frac{1}{|\mathcal{T}|} \text{ für } N \in \mathcal{T}$$

Posteriori-Verteilung im Capture-Recapture-Experiment

Berechne Posteriori-Wahrscheinlichkeitsfunktion:

$$f(N|x) = \frac{f(x|N)f(N)}{f(x)} = \frac{f(x|N)f(N)}{\sum_N f(x|N)f(N)}$$

Beachte: Erst **nach** der Beobachtung $X = x$ weiss man, dass mindestens $M + n - x$ Fische im See schwimmen, da die Likelihood $f(x|N) = 0$ ist für $x < M + n - N$, also für $N < M + n - x$. Weiterhin muss $n \leq N$ gelten.

Also hat die Posteriori-Verteilung $f(\theta|x)$ den Träger

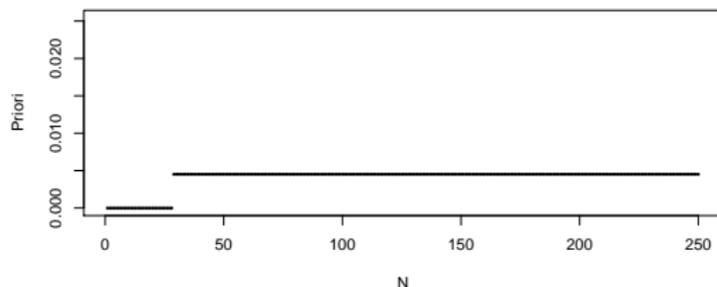
$$\mathcal{T} = \{\max(M + n - x, n), \max(M + n - x, n) + 1, \dots, Y_{\max}\}$$

Graphische Darstellung der Posteriori-Verteilung

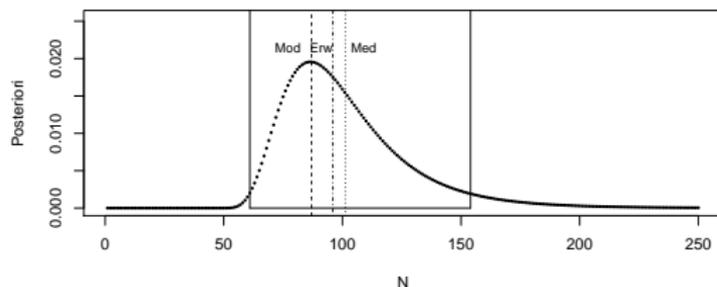
- Priori-Verteilung:
Gleichverteilung
- Berechnet:
 - ▶ Modus,
 - ▶ Median und
 - ▶ Erwartungswertder Posteriori-Verteilung und die HPD-Region, die hier ein Intervall ist.

Beispiel

Priori Verteilung



Posteriori Verteilung



Beispiel für $M = 29$, $x = 10$,
 $n = 30$ mit $Y_{\max} = 250$.

Bei der Posteriori-Verteilung
werden die 3 Punktschätzer
durch Linien verdeutlicht.
Die 95%-HPD-Region stellen
die äußeren Linien dar.

5.4 Modellanpassung

Häufig ist es von Interesse die Anpassung eines bestimmten stochastischen Modells an vorliegende Daten zu studieren. Dies ist insbesondere bei kategorialen Daten der Fall.

Beispiele:

1. Ist eine Population im Hardy-Weinberg-Gleichgewicht?

Untersuche N Individuen und berechne die empirischen Häufigkeiten der Genotypen aa , ab , bb . Die beobachtete Genotypverteilung ist z.B. $\mathbf{x} = (600, 320, 80)$.

Fortsetzung der Beispiele

2. Sind zwei Variablen unabhängig?

Hier wird untersucht, ob es eine Abhängigkeit zwischen Geschlecht und Promotionsinteresse gibt. Die vorliegenden Daten sind:

	Interesse		
	Ja	Nein	
♂	5	12	17
♀	6	5	11
	11	17	28

3. Im 19. Jahrhundert wurden in Sachsen Daten zur Häufigkeit von männlichen Nachkommen bei 6115 Familien mit jeweils (!) 12 Kindern erhoben:

# Jungen	0	1	2	3	4	5	6	7	8	9	10	11	12
# Familien	3	24	104	286	670	1033	1343	1112	829	478	181	45	7

Folgt die Verteilung einer Binomialverteilung?

Das saturierte Modell

In allen drei Beispielen möchte man ein bestimmtes Modell (“Null-Modell”, “Null-Hypothese”) mit dem allgemeinen Modell (“saturiertes” Modell) unter der Annahme einer Multinomialverteilung $\mathcal{M}_p(n, \pi)$ vergleichen.

Beispiel	Null-Modell	Anzahl der Parameter p	Anzahl der Kategorien K
1	Population ist im H-W Gleichgewicht	1	3
2	Variablen sind unabhängig	2	4
3	Daten sind binomial verteilt	1	13

Test auf Modellanpassung

Zum statistischen Testen der “Null-Hypothese” (H_0) geht man nach folgendem Schema vor:

- *ML*-Schätzung der unbekannt Parameter im Null-Modell

Beispiel 1: $\hat{q} = \frac{600 + \frac{320}{2}}{1000} = 0.76$

Beispiel 2: $\hat{\pi}_{\sigma} = \frac{17}{28}$

$$\hat{\pi}_{\text{Interesse}} = \frac{11}{28}$$

Beispiel 3: $\hat{\pi} = \frac{0 \cdot 3 + 1 \cdot 24 + \dots + 12 \cdot 7}{6115 \cdot 12} = 0.519215$

- Berechnung der erwarteten Anzahl E_i an Fällen in Kategorie i unter Annahme des Null-Modells

Test auf Modellanpassung II

- Berechnung des **Pearsonschen χ^2 -Maßes**

$$\chi^2 = \sum_{i=1}^K r_i^2 = \sum_{i=1}^K \frac{(X_i - E_i)^2}{E_i}$$

als Gesamtmaß für die Abweichung, wobei X_i die tatsächlich beobachteten Anzahlen in Kategorie i sind

Unter der Annahme, dass H_0 wahr ist, hat χ^2 eine (asymptotische) χ^2 -Verteilung mit $k = K - 1 - p$ Freiheitsgraden (mit K Anzahl der Kategorien und p Anzahl der Modellparameter).

Ermittlung des **p-Wertes**:

Wahrscheinlichkeit unter H_0 ein solches oder noch extremeres Resultat zu beobachten.

→ Berechnung über Quantile der χ^2 -Verteilung mit k Freiheitsgraden

Bemerkungen

- Die χ^2 -Verteilung gilt nur asymptotisch (für $n \rightarrow \infty$).
Faustregel: Es muss gelten, dass alle $E_j > 1$ und mindestens 80% der $E_j > 5$ sind.
- Alternativ bietet sich auch die Berechnung der **Devianz D** an:

$$D = 2 \cdot \sum_{i=1}^K X_i \log \left(\frac{X_i}{E_i} \right)$$

Diese besitzt unter H_0 die gleiche Verteilung wie χ^2 .

- Unter H_0 sind die χ^2 -Residuen r_i approximativ und asymptotisch standardnormalverteilt, d.h. Residuen mit $|r_i| > 2$ deuten auf schlechte Modellanpassung hin.

Berechnung der erwarteten Anzahlen in Beispiel 1

Unter Annahme des Hardy-Weinberg-Gleichgewichts erhält man:

$$\hat{\pi}_1 = \hat{q}^2 = 0.5776$$

$$\hat{\pi}_2 = 2\hat{q}(1 - \hat{q}) = 0.3648$$

$$\hat{\pi}_3 = (1 - \hat{q})^2 = 0.0576$$

Daher ergeben sich als erwartete Anzahlen bei $n = 1000$ Individuen:

$$E_1 = n \cdot \hat{\pi}_1 = 577.6$$

$$E_2 = n \cdot \hat{\pi}_2 = 364.8$$

$$E_3 = n \cdot \hat{\pi}_3 = 57.6$$

Berechnung der erwarteten Anzahlen in Beispiel 2

Unter Unabhängigkeit gilt z.B. für den Eintrag (♂, Promotionsinteresse = Ja):

$$\hat{\pi}_1 = \hat{\pi}_{\text{♂}} \cdot \hat{\pi}_{\text{Interesse}} = \frac{17}{28} \cdot \frac{11}{28}$$

Für die erwartete Anzahl folgt:

$$E_1 = n \cdot \hat{\pi}_1 = 28 \cdot \frac{17}{28} \cdot \frac{11}{28} = \frac{17 \cdot 11}{28} \approx 6.68$$

Die Werte der anderen Fälle erhält man analog (erwartete Anzahl in Klammern):

	Ja	Nein	
♂	5 (6.68)	12 (10.32)	17
♀	6 (4.32)	5 (6.68)	11
	11	17	28

Berechnung der erwarteten Anzahlen in Beispiel 3

Hier ergeben sich unter Verwendung der Wahrscheinlichkeit

$$P(X = x) \text{ mit } x = 0, \dots, 12$$

bei Vorliegen einer Binomialverteilung mit $n = 12$ und $\pi = 0.519215$ folgende erwartete Häufigkeiten:

$$E_i = P(X = i) \cdot 6115$$

Tabelle mit beobachteten und erwarteten Anzahlen:

	0	1	2	3	...	11	12
X_i	3	24	104	286	...	45	7
E_i	0.9	12.1	71.8	258.5	...	26.1	2.3

Vergleich von χ^2 und Devianz in den 3 Beispielen

Bsp.	χ^2	K	p	k	p -Wert	D	p -Wert
1	15,08	3	1	1	0,0001	14,36	0,00015
2	1,769	4	2	1	0,18	1,765	0,18
3	110,5	13	1	11	0	97,0	$6,66 \cdot 10^{-16}$