Vorlesung Statistik II

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Sommersemester 2026

Pivot-Variable

Wesentliche Voraussetzung für die dargestellte Konstruktion eines Konfidenzintervalls für den Parameter ϑ ist die Existenz einer Variable

- die vom Parameter ϑ abhängt
- deren Verteilung bekannt ist
- deren Verteilung nicht vom Parameter ϑ abhängt

Eine derartige Variable heißt *Pivot-Variable* und war im oben betrachteten Fall eines Konfidenzintervalls für den Erwartungswert μ bei bekannter Varianz σ^2 gegeben durch die $N(0,1)$-verteilte Standardisierte

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Konfidenzintervalle für den Erwartungswert μ einer normalverteilten Zufallsvariable bei unbekannter Varianz σ^2

Wir betrachten nun den Fall, dass bei einer normalverteilten Population neben dem Erwartungswert μ auch die Varianz σ^2 unbekannt ist

Der Parameter σ^2 ist dann ebenfalls aus den Daten zu schätzen

Als Schätzer für den Parameter σ^2 bietet sich die erwartungstreue Schätzfunktion der Stichprobenvarianz an

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

Verteilung der Pivot-Variable

Ebenso naheliegend erscheint dann die Verwendung der Pivot-Variable

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Dabei haben wir aber die feste Zahl σ durch die Zufallsvariable S ersetzt

Die Verteilung der Pivot-Variable wird durch die Verteilung der Zufallsvariable S natürlich mitbestimmt, die damit nicht mehr normalverteilt ist

Die Ableitung der Verteilung der Pivot-Variable gründet sich auf folgende Eigenschaften

Verteilung der Pivot-Variable

Sind die Zufallsvariablen $X_1, \ldots, X_n$ stochastisch unabhängig und identisch normalverteilt gemäß $N(\mu, \sigma^2)$, dann ist

- das arithmetische Mittel

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

  normalverteilt gemäß $N(\mu, \sigma^2/n)$

- die "normierte" Stichprobenvarianz

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

  Chi-Quadrat-verteilt mit $n-1$ Freiheitsgraden

Verteilung der Pivot-Variable

Die Schätzfunktionen $\bar{X}$ für den Erwartungswert μ und S^2 für die Varianz σ^2 sind stochastisch unabhängig genau dann, wenn die (stochastisch unabhängigen) Zufallsvariablen $X_1, \ldots, X_n$ identisch normalverteilt sind

Diese Eigenschaft

- zeichnet die Normalverteilung vor allen anderen Verteilungen aus
- ist fundamental für die Ableitung der Verteilung der Pivot-Variable

$$\frac{\bar{X} - \mu}{S/\sqrt{n}},$$

da nur dann die Verteilung der Variable im Zähler unabhängig von der Verteilung der Variable im Nenner ist

Verteilung der Pivot-Variable

Bei der Besprechung der Testverteilungen hatten wir folgenden Fall betrachtet

- Seien die Zufallsvariablen Z $N(0,1)$-verteilt und Y $\chi^2(n)$-verteilt, sowie Z und Y stochastisch unabhängig
- Dann ist die Zufallsvariable

$$T = \frac{Z}{\sqrt{Y/n}}$$

  t-verteilt mit n Freiheitsgraden, kurz $t(n)$-verteilt
- Diese Tatsache können wir im vorliegenden Fall nutzen

Verteilung der Pivot-Variable

Die Zufallsvariable $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ ist $N(0,1)$-verteilt

Die hiervon stochastisch unabhängige Zufallsvariable

$$Y = \frac{(n-1)\,S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

ist $\chi^2(n-1)$-verteilt

Dann ist die Zufallsvariable

$$\frac{Z}{\sqrt{Y/(n-1)}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)\,S^2}{\sigma^2\,(n-1)}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

t-verteilt mit $n-1$ Freiheitsgraden

Damit haben wir die Verteilung der oben betrachteten Pivot-Variable als t-Verteilung mit $n-1$ Freiheitsgraden identifiziert

$(1-\alpha)$-Konfidenzintervall

Für die Pivot-Variable lässt sich dann unmittelbar ein zweiseitig beschränkter Bereich angeben, in dem sie mit Wahrscheinlichkeit $1-\alpha$ liegt

$$P\left(-t_{1-\alpha/2}(n-1) \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{1-\alpha/2}(n-1)\right) = 1-\alpha,$$

mit $t_{1-\alpha/2}(n-1)$ das $(1-\alpha/2)$-Quantil der $t(n-1)$-Verteilung

Damit erhält man ganz analog zum Fall mit bekannter Varianz ein $(1-\alpha)$-Konfidenzintervall durch

$$\left[\bar{X} - t_{1-\alpha/2}(n-1)\cdot\frac{S}{\sqrt{n}},\ \bar{X} + t_{1-\alpha/2}(n-1)\cdot\frac{S}{\sqrt{n}}\right]$$

Beispiel

Wir betrachten wieder das oben beschriebene Experiment zur Tonhöhenwahrnehmung, in dem der Punkt der subjektiven Gleichheit zu einem Sinuston von 1000 Hz mit Hilfe der Herstellungsmethode ermittelt wird

In diesem Fall nehmen wir an, dass die Standardabweichung nicht bekannt ist, dass sich bei einer Schätzung über die Stichprobenvarianz jedoch ebenfalls der Wert $s = 7.5\,\text{Hz}$ ergibt

Der Mittelwert der Einstellungen sei wiederum $\bar{x} = 996\,\text{Hz}$

Beispiel

Ein Konfidenzintervall zum Niveau 0.95 erhält man dann auf folgende Weise

- Mit Hilfe einer Tabelle der t-Verteilung bei $30-1$ Freiheitsgraden oder mit R durch Eingabe von `qt(0.975, 29)` ermittelt man für das 0.975-Quantil den Wert $t_{0.975}(29) = 2.05$
- Mit $t_{0.975}(29) \cdot \frac{s}{\sqrt{n}} = 2.05 \cdot \frac{7.5}{\sqrt{30}} = 2.8$ ergibt sich das 0.95-Konfidenzintervall

$$\begin{aligned}
&\left[\bar{x} - t_{1-\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}\right] \\
&= [996 - 2.8, 996 + 2.8] \\
&= [993.2, 998.8]
\end{aligned}$$

Einseitiges $(1-\alpha)$-Konfidenzintervall

Analog zur Situation bei bekannter Varianz ergeben sich die einseitigen Konfidenzintervalle zum Konfidenzniveau $1-\alpha$ durch

$$\left[\bar{X} - t_{1-\alpha}(n-1) \cdot \frac{S}{\sqrt{n}},\ \infty\right)$$

für die Begrenzung nach unten und

$$\left(-\infty,\ \bar{X} + t_{1-\alpha}(n-1) \cdot \frac{S}{\sqrt{n}}\right]$$

für die Begrenzung nach oben

Konfidenzintervalle für den Erwartungswert μ einer beliebig verteilten Zufallsvariable bei großem Stichprobenumfang ($n > 30$)

Mit dem zentralen Grenzwertsatz kann die Zufallsvariable $\bar{X}$ als approximativ normalverteilt angesehen werden

Approximatives zweiseitiges $(1-\alpha)$-Konfidenzintervall bei bekannter Varianz σ^2

$$\left[\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Approximatives zweiseitiges $(1-\alpha)$-Konfidenzintervall bei unbekannter Varianz σ^2

$$\left[\bar{X} - t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}},\ \bar{X} + t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}}\right]$$

Konfidenzintervalle für die Varianz σ^2 einer normalverteilten Zufallsvariable

Hierbei gehen wir wieder von der Tatsache aus, dass die "normierte" Stichprobenvarianz für stochastisch unabhängige und $N(\mu, \sigma^2)$-verteilte Zufallsvariablen

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

$\chi^2(n-1)$-verteilt ist

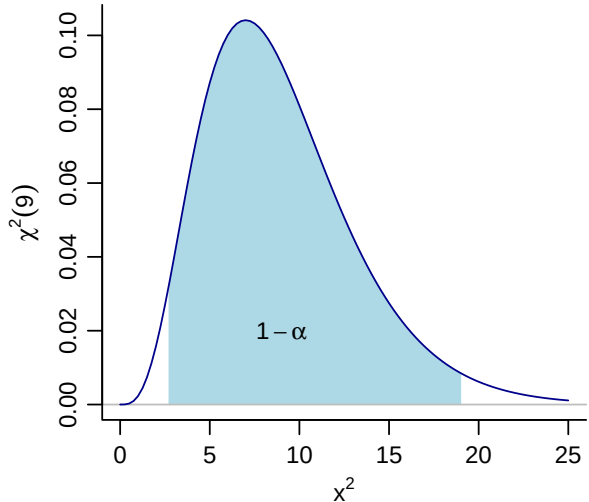
Für ein zweiseitiges Konfidenzintervall und vorgegebene Irrtumswahrscheinlichkeit α betrachten wir dann das $(\alpha/2)$-Quantil und das $(1-\alpha/2)$-Quantil

$$q_{\alpha/2} = \chi^2_{\alpha/2}(n-1) \quad \text{und} \quad q_{1-\alpha/2} = \chi^2_{1-\alpha/2}(n-1)$$

der Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden

Dichte der χ^2-Verteilung

Dichte der χ^2-Verteilung mit 9 Freiheitsgraden



$(1-\alpha)$-Konfidenzintervall

Dann hat man

$$\begin{aligned} 1-\alpha &= P\left(q_{\alpha/2} \leq \frac{(n-1)\,S^2}{\sigma^2} \leq q_{1-\alpha/2}\right) \\ &= P\left(\frac{q_{\alpha/2}}{(n-1)\,S^2} \leq \frac{1}{\sigma^2} \leq \frac{q_{1-\alpha/2}}{(n-1)\,S^2}\right) \\ &= P\left(\frac{(n-1)\,S^2}{q_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)\,S^2}{q_{\alpha/2}}\right) \end{aligned}$$

Daraus erhält man dann ein Konfidenzintervall zum Niveau $1-\alpha$ durch

$$\left[\frac{(n-1)\,S^2}{q_{1-\alpha/2}},\ \frac{(n-1)\,S^2}{q_{\alpha/2}}\right]$$

Beispiel

In dem betrachteten Experiment zur Tonhöhenwahrnehmung ergebe sich als Schätzung von σ^2 über die Stichprobenvarianz der Wert $s^2 = 7.5^2 = 56.25\,\text{Hz}^2$

Für die vorgegebene Irrtumswahrscheinlichkeit $\alpha = 0.05$ lauten die Quantile der $\chi^2(n-1)$-Verteilung dann

$$q_{\alpha/2} = q_{0.025} = 16.05, \quad q_{1-\alpha/2} = q_{0.975} = 45.72$$

Damit erhält man als 0.95-Konfidenzintervall

$$\left[\frac{(n-1)\,s^2}{q_{1-\alpha/2}},\ \frac{(n-1)\,s^2}{q_{\alpha/2}}\right] = \left[\frac{29 \cdot 56.25}{45.72},\ \frac{29 \cdot 56.25}{16.05}\right]$$
$$= [35.68,\ 101.65]$$

Durch Wurzelziehen ergibt sich daraus das 0.95-Konfidenzintervall für σ $[6.0,\ 10.1]\,\text{Hz}$

Konfidenzintervalle für die Wahrscheinlichkeit p einer Bernoulli-verteilten Zufallsvariable

Die binären Zufallsvariablen $X_1, \ldots, X_n$ seien stochastisch unabhängig und identisch Bernoulli-verteilt mit dem Parameter p

Dann ist die Summenvariable $\sum_{i=1}^{n} X_i$ binomialvereilt mit Erwartungswert $E\left(\sum_{i=1}^{n} X_i\right) = n\,p$ und Varianz $Var\left(\sum_{i=1}^{n} X_i\right) = n\,p\,(1-p)$

Für hinreichend großen Stichprobenumfang n (es genügt $n\,p\,(1-p) > 9$) ist dann das arithmetische Mittel $\bar{X} = \sum_{i=1}^{n} X_i/n$ nach dem zentralen Grenzwertsatz approximativ normalverteilt, sodass gilt

$$\frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} \sim N(0,1)$$

$(1-\alpha)$-Konfidenzintervall

Wegen $E(\bar{X}) = p$ und Varianz $Var(\bar{X}) = p\,(1-p)\,/\,n$ gilt damit approximativ

$$\frac{\bar{X} - p}{\sqrt{p\,(1-p)\,/\,n}} \sim N(0,1)$$

Wird nun im Nenner zusätzlich der Parameter p durch die relative Häufigkeit $\hat{p} = \bar{X}$ approximiert, dann ergibt sich für vorgegebene Irrtumswahrscheinlichkeit α

$$1 - \alpha \approx P\left(-z_{1-\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}\,(1-\hat{p})\,/\,n}} \leq z_{1-\alpha/2}\right)$$

$(1-\alpha)$-Konfidenzintervall

Analog zu den bereits betrachteten Fällen hat man folgende Umformung

$$1-\alpha \approx P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right)$$

Als zweiseitiges Konfidenzintervall zum Niveau $1-\alpha$ erhält man dann

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}},\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right]$$

Beispiel: Geburtenrate

Unter den 87827 Lebendgeburten in Niedersachsen im Jahre 1972 waren 45195 Jungen

Als Konfidenzintervall zum Niveau 0.99 für die Wahrscheinlichkeit p, dass ein Junge geboren wird, ergibt sich mit $z_{0.995} = 2.576$ und

$$\hat{p} = \frac{45195}{87827} = 0.515 \quad \text{und}$$
$$\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = \sqrt{\frac{0.515\,(1-0.515)}{87827}} = 0.0017$$

dann

$$[0.515 - 2.576 \cdot 0.0017,\ 0.515 + 2.576 \cdot 0.0017] = [0.511,\ 0.519]$$

Beispiel: Stichprobenumfang

Vor einer Bundestagswahl möchte ein Meinungsforschungsinstitut eine Prognose über den prozentualen Stimmenanteil abgeben, den eine Partei in dieser Wahl erreichen wird

Wie viele zufällig ausgewählte Wahlberechtigte müssen mindestens befragt werden, um ein Konfidenzintervall zum Niveau 0.95 zu erhalten, dessen Breite höchstens 2% beträgt?

Es sei also

$$2 \cdot z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \leq 0.02$$

Beispiel: Stichprobenumfang

Dann gilt

$$n \geq \frac{4 \cdot z_{1-\alpha/2}^2 \cdot \hat{p} \cdot (1 - \hat{p})}{0.02^2}$$

Der Stichprobenumfang hängt also über den Faktor $\hat{p} \cdot (1 - \hat{p})$ von der relativen Häufigkeit ab, wobei der Faktor für $\hat{p} = 0.5$ maximal wird

Mit diesem Wert ergibt sich als Abschätzung

$$n \geq \frac{4 \cdot 1.96^2 \cdot 0.5 \cdot 0.5}{0.0004} = 9603.65$$

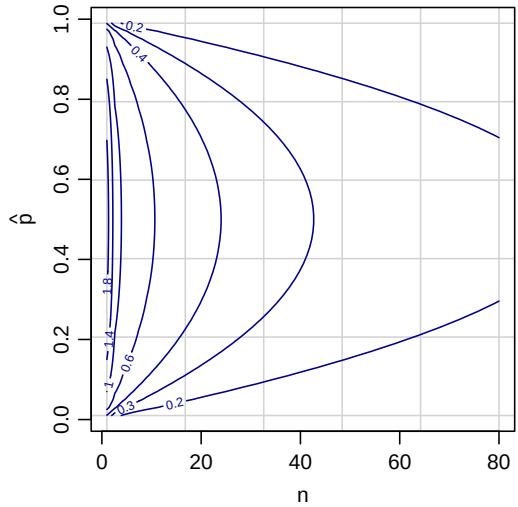
Es müssen also mindestens 9604 Wahlberechtigte befragt werden

Breite der 0.95-Konfidenzintervalle

Breite in Abhängigkeit von der relativen Häufigkeit $\hat{p}$ und dem Stichprobenumfang n

```
alpha <- 0.05
phat  <- seq(0.01, 0.99, length = 50)
n     <- seq(1, 80, length = 50)

## Linien gleicher Breite
contour(n, phat,
  outer(n, phat,
    function(x, y)
      2*qnorm(1 - alpha/2) *
      sqrt(y*(1 - y)/x)),
  xlab = "n",
  ylab = expression(hat(p)))
```



Konfidenzintervalle für den Erwartungswert μ einer normalverteilten Zufallsvariable

Zweiseitiges $(1-\alpha)$-Konfidenzintervall bei bekannter Varianz σ^2

$$\left[\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Zweiseitiges $(1-\alpha)$-Konfidenzintervall bei unbekannter Varianz σ^2

$$\left[\bar{X} - t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}},\ \bar{X} + t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}}\right]$$

Konfidenzintervalle für den Erwartungswert μ einer normalverteilten Zufallsvariable

Einseitige $(1-\alpha)$-Konfidenzintervalle bei bekannter Varianz σ^2

$$\left[\bar{X} - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}},\ \infty\right) \quad \text{und} \quad \left(-\infty,\ \bar{X} + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Einseitige $(1-\alpha)$-Konfidenzintervalle bei unbekannter Varianz σ^2

$$\left[\bar{X} - t_{1-\alpha}(n-1) \cdot \frac{S}{\sqrt{n}},\ \infty\right) \quad \text{und} \quad \left(-\infty,\ \bar{X} + t_{1-\alpha}(n-1) \cdot \frac{S}{\sqrt{n}}\right]$$

Konfidenzintervalle für den Erwartungswert μ einer beliebig verteilten Zufallsvariable bei großem Stichprobenumfang ($n > 30$)

Approximatives zweiseitiges $(1-\alpha)$-Konfidenzintervall bei bekannter Varianz σ^2

$$\left[\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Approximatives zweiseitiges $(1-\alpha)$-Konfidenzintervall bei unbekannter Varianz σ^2

$$\left[\bar{X} - t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}},\ \bar{X} + t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}}\right]$$

Konfidenzintervalle für den Erwartungswert μ einer beliebig verteilten Zufallsvariable bei großem Stichprobenumfang ($n > 30$)

Approximative einseitige $(1-\alpha)$-Konfidenzintervalle bei bekannter Varianz σ^2

$$\left[\bar{X} - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}},\ \infty\right) \quad \text{und} \quad \left(-\infty,\ \bar{X} + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Approximative einseitige $(1-\alpha)$-Konfidenzintervalle bei unbekannter Varianz σ^2

$$\left[\bar{X} - t_{1-\alpha}(n-1) \cdot \frac{S}{\sqrt{n}},\ \infty\right) \quad \text{und} \quad \left(-\infty,\ \bar{X} + t_{1-\alpha}(n-1) \cdot \frac{S}{\sqrt{n}}\right]$$

Konfidenzintervalle für die Varianz σ^2 einer normalverteilten Zufallsvariable

Zweiseitiges $(1-\alpha)$-Konfidenzintervall für σ^2

$$\left[\frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}(n-1)},\ \frac{(n-1)\,S^2}{\chi^2_{\alpha/2}(n-1)}\right]$$

Einseitige $(1-\alpha)$-Konfidenzintervalle für σ^2

$$\left[\frac{(n-1)\,S^2}{\chi^2_{1-\alpha}(n-1)},\ \infty\right) \quad \text{und} \quad \left(-\infty,\ \frac{(n-1)\,S^2}{\chi^2_{\alpha}(n-1)}\right]$$

Konfidenzintervalle für die Wahrscheinlichkeit p einer Bernoulli-verteilten Zufallsvariable ($n \cdot \bar{X} \cdot (1 - \bar{X}) > 9$)

Zweiseitiges $(1 - \alpha)$-Konfidenzintervall für p

$$\left[\bar{X} - z_{1-\alpha/2} \sqrt{\frac{\bar{X}\,(1-\bar{X})}{n}},\ \bar{X} + z_{1-\alpha/2} \sqrt{\frac{\bar{X}\,(1-\bar{X})}{n}} \right]$$

Einseitige $(1 - \alpha)$-Konfidenzintervalle für p

$$\left[\bar{X} - z_{1-\alpha} \sqrt{\frac{\bar{X}\,(1-\bar{X})}{n}},\ \infty \right) \quad \text{und} \quad \left(-\infty,\ \bar{X} + z_{1-\alpha} \sqrt{\frac{\bar{X}\,(1-\bar{X})}{n}} \right]$$

Bayessche Konfidenzintervalle

Vorsicht bei der Interpretation "frequentistischer" Konfidenzintervalle!

Das Konfidenzniveau $1 - \alpha$ kennzeichnet

- *nicht* die Wahrscheinlichkeit, mit der der Parameter in einem resultierenden Intervall liegt
- die Wahrscheinlichkeit, mit der das verwendete Verfahren (bei wiederholter Anwendung auf mögliche Datensätze) ein Intervall liefert, in dem der Parameter enthalten ist

Für Konfidenzintervalle, die im Rahmen der Bayesschen Statistik abgeleitet werden, kann das Konfidenzniveau tatsächlich als die Wahrscheinlichkeit interpretiert werden, mit der der Parameter im resultierenden Intervall liegt

Konstruktion Bayesscher Konfidenzintervalle

Die Konstruktion eines Bayesschen Konfidenzintervalls für den Parameter ϑ bezieht sich auf die A-posteriori-Verteilung bzw. deren Dichtefunktion $f(\vartheta \mid x_1, \ldots, x_n)$

Für ein zweiseitiges Konfidenzintervall zum Niveau $1 - \alpha$ bestimmt man

- die untere Grenze g_u des Intervalls als $(\alpha/2)$-Quantil der A-posteriori-Verteilung, d. h.
$$\int_{-\infty}^{g_u} f(\vartheta \mid x_1, \ldots, x_n) d\vartheta = \frac{\alpha}{2}$$
- die obere Grenze g_o des Intervalls als $(1 - \alpha/2)$-Quantil der A-posteriori-Verteilung, d. h.
$$\int_{-\infty}^{g_o} f(\vartheta \mid x_1, \ldots, x_n) d\vartheta = 1 - \frac{\alpha}{2}$$

Beispiel

Für die oben besprochene Anwendung der Konstanzmethode (20 Ja-Antworten, 30 Durchgänge) mit dem Hyperparameter $c = 2.0$ der A-priori-Verteilung ergibt sich das 0.95-Konfidenzintervall für p $[0.48, 0.80]$

