6 Intervallschätzer und Konfidenzbänder

Im Rahmen der Konzepte rund um das Schätzen haben wir bislang sogenannte *Punktschätzer* betrachtet, d. h., ein Schätzer S ‚trifft *ein* Element' aus seinem Bildraum, vgl. Kap. 4. Alternativ dazu befassen wir uns in diesem Kapitel mit sogenannten *Intervallschätzern.* Die Grundidee ist, dass wir ganze Intervalle schätzen, die den wahren unbekannten Parameter mit großer Wahrscheinlichkeit überdecken. Dies wird uns erste Einblicke in die Ideen der Hypothesentests liefern, auch wenn wir die klassischen Begrifflichkeiten des Testens noch nicht einführen.

6.1 Definition und einfache Eigenschaften

Es sei $\mathbf{x} = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ ein Beobachtungsvektor. Wir betrachten ein Modell bestehend aus einem Zufallsvektor $\mathfrak{X} = (X_1, \ldots, X_n)^t$ und einer Verteilungsfamilie $(\nu_\vartheta)_{\vartheta \in \Theta}$, und interessieren uns für eine reellwertige abgeleitete Kenngröße der Verteilung. Beispielsweise könnte man sich im Falle des Modells gegeben durch unabhängige und identisch verteilte Zufallsvariable $X_1, \ldots, X_n$ mit $X_1 \sim \nu_\vartheta$, wobei ν_ϑ Mitglied der Familie $(\nu_\vartheta)_{\vartheta \in \Theta}$ aller integrierbaren Verteilungen sei, für den Erwartungswert $\mu_\vartheta := \mathbb{E}_\vartheta[X_1]$ interessieren.

Die in diesem Abschnitt betrachteten Statistiken sind sogenannte *Konfidenzintervalle,* bezeichnet durch I, bestehend aus einer unteren Intervallgrenze $T^{(u)}(\mathbf{x})$ und einer oberen Intervallgrenze $T^{(o)}(\mathbf{x})$, genauer

$$I(\mathbf{x}) := [T^{(u)}(\mathbf{x}), T^{(o)}(\mathbf{x})] \subseteq \mathbb{R},$$

mit $T^{(u)}(\mathbf{x}) \leq T^{(o)}(\mathbf{x})$, welche wir als reellwertige Statistiken verstehen, siehe auch Abb. 6.1. Wie immer unterscheiden wir zwischen dem beobachteten, festen Intervall $I(\mathbf{x})$ und dem zufälligen Intervall $I(\mathfrak{X})$.

M. Messer und G. Schneider, *Statistik,* https://doi.org/10.1007/978-3-662-59339-4_6

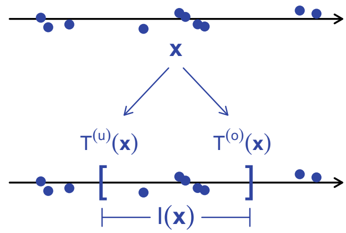


Abb. 6.1 Konfidenzintervall als Statistik $I(x)$

Definition 6.1 (Konfidenzintervall)
Es sei ein statistisches Modell gegeben durch einen Zufallsvektor $\mathfrak{X} = (X_1, \ldots, X_n)^t$ und eine Verteilungsfamilie $(\nu_\vartheta)_{\vartheta \in \Theta}$. Zudem sei $\tau(\vartheta)$ ein reellwertiger abgeleiteter Parameter und $\alpha \in (0, 1)$. Seien weiter $T^{(u)}(\mathbf{x})$ und $T^{(o)}(\mathbf{x})$ zwei reellwertige Statistiken mit $T^{(u)} \leq T^{(o)}$. Dann heißt ein Intervall I gegeben durch

$$I(\mathbf{x}) := \left[T^{(u)}(\mathbf{x}), T^{(o)}(\mathbf{x})\right]$$

ein $(1-\alpha)$-Konfidenzintervall für $\tau(\vartheta)$, falls für alle $\vartheta \in \Theta$ gilt

$$\mathbb{P}_\vartheta\left(I(\mathfrak{X}) \ni \tau(\vartheta)\right) \geq 1 - \alpha. \tag{6.1}$$

Ein $(1-\alpha)$-Konfidenzintervall nennen wir auch ein Konfidenzintervall *zum Niveau* $1-\alpha$.

Bedeutung des Konfidenzintervalls: Wenn ν_ϑ die wahre zugrunde liegende Verteilung ist, dann überdeckt das zufällige Intervall $I(\mathfrak{X})$ die abgeleitete Kenngröße $\tau(\vartheta)$ mit großer Wahrscheinlichkeit, genauer mit Wahrscheinlichkeit mindestens $1-\alpha$. Die Idee ist, dass α ‚klein' gewählt wird, denn wir wollen den wahren Parameter mit großer Wahrscheinlichkeit überdecken. In der Praxis wählt man α oft als 5 %, 1 % oder 0.1 %. Zu gegebenem α möchten wir außerdem das Intervall möglichst klein konstruieren, da ein großes Intervall weniger Informationen zur Schätzung von $\tau(\vartheta)$ liefert.

Wichtig ist hier: Der Zufall steckt nach Konstruktion im Intervall $I(\mathfrak{X})$. Der abgeleitete Parameter $\tau(\vartheta)$ ist wie auch der Parameter ϑ selbst eine feste Größe, d. h. nicht als zufällig modelliert. Daher sagen wir auch, dass das zufällige Intervall $I(\mathfrak{X})$ den Parameter $\tau(\vartheta)$ überdeckt, kurz $\{I(\mathfrak{X}) \ni \tau(\vartheta)\}$, und vermeiden die missverständliche Formulierung, dass der Parameter $\tau(\vartheta)$ in das Intervall $I(\mathfrak{X})$ fällt ($\cancel{\{\tau(\vartheta) \in I(\mathfrak{X})\}}$). Mathematisch sind diese beiden Ereignisse $\{I(\mathfrak{X}) \ni \tau(\vartheta)\}$ und $\{\tau(\vartheta) \in I(\mathfrak{X})\}$ gleich, aber letztere Schreibweise könnte fälschlicherweise vermuten lassen, dass der Zufall in $\tau(\vartheta)$ steckt.

Wir können auch von *asymptotischen* Konfidenzintervallen sprechen. Es sei dazu ein Modell mit Vektor $\mathfrak{X}_\infty = (X_1, X_2, \ldots)^t$ und Verteilungsfamilie $(\nu_\vartheta)_{\vartheta \in \Theta}$ gegeben. Wir betrachten dann für $n = 1, 2, \ldots$ das Modell, welches durch die Restriktion auf die ersten n Komponenten ($\mathfrak{X}_n = (X_1, \ldots, X_n)^t$) hervorgeht, sowie jeweils Statistiken $T_n^{(u)}$ und $T_n^{(o)}$, sodass wir eine Folge von Konfidenzintervallen $(I_n)_{n=1,2,\ldots}$ erhalten. Wir fordern dann, dass die Überdeckungswahrscheinlichkeit (6.1) asymptotisch für $n \to \infty$ gilt, d. h.

$$\lim_{n\to\infty} \mathbb{P}_\vartheta \left(I_n(\mathfrak{X}_n) \ni \tau(\vartheta)\right) \geq 1 - \alpha.$$

Wir sprechen dann auch kurz von dem asymptotischen Konfidenzintervall I_n.

6.2 Interpretation und Formulierung

Für ein Konfidenzintervall konstruiert man also Intervallgrenzen so, dass die Forderung an die Überdeckungswahrscheinlichkeit (6.1) eingehalten wird. Wir betrachten zunächst ein prominentes Beispiel eines Konfidenzintervalls für den Erwartungswert normalverteilter Beobachtungen. Im Anschluss diskutieren wir dessen Bedeutung – das wird ein wichtiger Schritt des Erlernens statistischer Denkweisen sein.

Beispiel 6.2 (Konfidenzintervall für μ bei Normalverteilung und bekanntem σ^2)
Es seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariable mit $X_1 \sim N(\mu, \sigma^2)$ mit $\mu \in \Theta := \mathbb{R}$. Es sei $\sigma^2 \in \mathbb{R}^+$ bekannt. Weiter sei q das 97.5 %-Quantil der Standardnormalverteilung (d. h. $q \approx 1.96$). Dann ist

$$I(\mathbf{x}) := \left[\bar{x}_n - q \cdot \frac{\sigma}{\sqrt{n}}, \bar{x}_n + q \cdot \frac{\sigma}{\sqrt{n}}\right] \tag{6.2}$$

ein 95 %-Konfidenzintervall für μ. Um die Überdeckungswahrscheinlichkeit (6.1) einzusehen, erinnern wir uns daran, dass die Summe unabhängiger normalverteilter Zufallsvariablen wieder normalverteilt ist, und da sich bei Unabhängigkeit auch die Erwartungswerte und Varianzen addieren, ist $\sum X_i \sim N(n\mu, n\sigma^2)$ und nach Standardisieren $(\bar{X}_n - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$. Damit gilt

$$\begin{aligned}
0.95 &= \mathbb{P}_\mu \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \in [-q, q]\right) \\
&= \mathbb{P}_\mu \left(\mu - q\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n \leq \mu + q\frac{\sigma}{\sqrt{n}}\right) \\
&= \mathbb{P}_\mu \left(\bar{X}_n + q\frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X}_n - q\frac{\sigma}{\sqrt{n}}\right) \\
&= \mathbb{P}_\mu(I(\mathfrak{X}) \ni \mu).
\end{aligned}$$

Wir bemerken, dass für die Überdeckungswahrscheinlichkeit (6.1) hier sogar Gleichheit gilt. Für ein allgemeines $(1-\alpha)$-Konfidenzintervall brauchen wir nur das 97.5 %-Quantil durch das $(1-\alpha/2)$-Quantil der $N(0, 1)$-Verteilung zu ersetzen.

Anhand von Abb. 6.2 diskutieren wir nun die Bedeutung des zufälligen Intervalls $I(\mathfrak{X})$. Dann diskutieren wir die Unterschiede zu dem auf den Beobachtungen $\mathbf{x}$ basierenden Konfidenzintervall $I(\mathbf{x})$. Wie auch vorher interpretieren wir $I(\mathbf{x})$ als Realisierung von $I(\mathfrak{X})$.

1. Bedeutung von $I(\mathfrak{X})$: Abb. 6.2 zeigt die Normalverteilungsdichte (rot) mit $\mu = 80$ und Standardabweichung $\sigma = 15$. Realisieren wir Stichproben der Größe $n = 50$ aus dieser Verteilung und bilden für jede Realisierung ein entsprechendes 95 %-Konfidenzintervall, so könnten wir zum Beispiel die in Schwarz bzw. Rosa dargestellten Intervalle erhalten. Nach Konstruktion erwarten wir bei 100 Stichproben, dass im Mittel fünf dieser Intervalle das wahre μ nicht überdecken. In Abb. 6.2 waren es durch Zufall vier Intervalle (rosa). Anhand dieser Realisierungen bekommen wir ein gutes Gefühl dafür, wie das zufällige Konfidenzintervall verteilt ist, wenn $\mu = 80$. Wir wiederholen die Aussage der Überdeckungswahrscheinlichkeit: Wenn die Beobachtungen aus der $N(80, 15^2)$-Verteilung stammen, überdeckt das *zufällige* Konfidenzintervall $I(\mathfrak{X})$ den wahren Parameter $\mu = 80$ mit Wahrscheinlichkeit 95 %.
2. Bedeutung von $I(\mathbf{x})$: Nun betrachten wir 50 Beobachtungen $\mathbf{x} = (x_1, \ldots, x_{50})^t$ (blaue Punkte) und ihr analog konstruiertes 95 %-KI $I(\mathbf{x})$ (blau) um den empirischen Mittelwert (grün). Wir stellen fest: Es überdeckt den theoretischen Wert $\mu = 80$ nicht! Es wäre damit in Punkt 1. rosa markiert, genau wie die Intervalle, die durch Zufall nur mit Wahrscheinlichkeit 5 % auftreten. Wenn also $\mu = 80$ wahr wäre, d. h., wenn die blauen Beobachtungen aus der roten Verteilung stammten, dann wäre etwas Unwahrscheinliches eingetreten. In diesem Sinne sind die Beobachtungen nicht mit der roten Verteilung, d. h. mit der Annahme $\mu = 80$, verträglich.
   Besonders interessant ist am Konfidenzintervall, dass offenbar dieselbe Interpretation für jeden Wert von μ außerhalb von $I(\mathbf{x})$ gilt. Wir können mithilfe des Konfidenzintervalls also mit einfachen Mitteln schnell einen Überblick bekommen, mit welchen Annahmen die Beobachtungen mehr und mit welchen sie weniger gut verträglich sind.

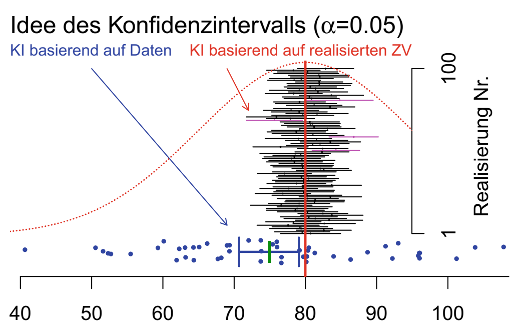


Abb. 6.2 Interpretation des Konfidenzintervalls

Zusätzlich gibt uns auch die Entfernung von μ eine Intuition für den Grad der Verträglichkeit: Betrachten wir ein $\tilde{\mu}$ weit außerhalb des 95 %-Konfidenzintervalls (zum Beispiel $\tilde{\mu} = 100$), so bekommen wir sofort einen Eindruck, dass dieses $\tilde{\mu}$ auch für kleinere α (vielleicht sogar $\alpha = 0.01$ oder 0.001) außerhalb des entsprechenden $(1-\alpha)$-Konfidenzintervalls liegen würde. Denn ein größeres Konfidenzintervall entspricht einem größeren Quantil q und entsprechend kleinerem α. Mit einem Wert weit außerhalb des 95 %-Konfidenzintervalls sind die Beobachtungen also sehr schlecht verträglich.

3. Fehlinterpretation von $I(\mathbf{x})$: Aus den Punkten 1. und 2. würden wir für die Praxis, in der wir ja ein konkretes Intervall $I(\mathbf{x})$ (zum Beispiel in Abb. 6.2 ca. $[71, 79]$) gegeben haben, nur zu gern eine Wahrscheinlichkeitsaussage herleiten. Zum Beispiel würden wir sehr gerne sagen: ‚μ liegt mit Wahrscheinlichkeit 95 % in $[71, 79]$'. Oder vielleicht besser: ‚$[71, 79]$ überdeckt μ mit Wahrscheinlichkeit 95 %' ? Diese Aussagen klingen kaum anders als die in 2., sind aber leider in dieser Form nicht möglich. Denn das auf den Beobachtungen basierende Konfidenzintervall $I(\mathbf{x})$ ist interpretiert als (feste) Realisierung; der Zufall ist bereits eingetreten, daher kann man dafür keine Wahrscheinlichkeitsaussage mehr machen. Das Intervall ist fest, und μ ist fest. Entweder das Intervall überdeckt den Parameter, oder es überdeckt ihn nicht. Wir wissen es leider nicht. Wem das spanisch vorkommt, der bedenke: Bei einem fairen Münzwurf, bei dem der Ausgang Kopf eingetreten ist, folgern wir auch nicht, dass der Ausgang des Experiments mit Wahrscheinlichkeit $1/2$ Kopf und mit Wahrscheinlichkeit 1/2 Zahl ist. Denn der Ausgang ist Kopf mit Wahrscheinlichkeit 1 – oder besser: ‚Die Münze zeigt Kopf'.
   Um Missverständnissen vorzubeugen, kann man sagen: ‚Ein auf den Beobachtungen x basierendes 95 %-Konfidenzintervall für μ ist $[71, 78]$.'

6.3 Ein asymptotisches Konfidenzintervall für den Erwartungswert

Wir wenden uns wieder der Theorie zu und betrachten nun *asymptotische* Konfidenzintervalle am Beispiel des Erwartungswerts. Ziel dieses Abschnittes ist es, Beispiel 6.2 mit elementaren stochastischen Mitteln zu verallgemeinern. Erstens möchten wir von der Normalverteilungsannahme weggehen und auch in nichtparametrischen Modellen ein Konfidenzintervall für den Erwartungswert angeben. Zweitens soll der Parameter σ, die Variabilität der Beobachtungen, die in der Praxis nicht als bekannt angenommen werden kann, durch einen geeigneten Schätzer approximiert werden.

Das Tolle ist, dass das entsprechende Konfidenzintervall die gleiche Struktur haben wird wie das Konfidenzintervall in (6.2). Wir werden sehen, dass im Rahmen nichtparametrischer Modelle unter gewissen Bedingungen

$$I_n(\mathbf{x}_n) := \left[\bar{x}_n - q \cdot \frac{s_n(\mathbf{x}_n)}{\sqrt{n}}, \bar{x}_n + q \cdot \frac{s_n(\mathbf{x}_n)}{\sqrt{n}}\right] \tag{6.3}$$

ein asymptotisches $(1-\alpha)$-Konfidenzintervall für den Erwartungswert der Individualbeobachtung ist, wobei q das $(1-\alpha/2)$-Quantil der Standardnormalverteilung bezeichnet. Im Vergleich zum Konfidenzintervall aus Beispiel 6.2 wurde hier die Standardabweichung σ durch den Schätzer $s_n(\mathbf{x}_n)$ ersetzt (vgl. (3.3)). Der wesentliche Grund für die Allgemeingültigkeit des Konfidenzintervalls (6.3) ist der Zentrale Grenzwertsatz: Der Mittelwert ist asymptotisch normalverteilt!

Konfidenzintervall für den Erwartungswert bei bekannter Varianz Wir betrachten zur Verallgemeinerung im Folgenden den Fall, dass $X_1, X_2, \dots$ unabhängige und identisch verteilte Zufallsvariable beschreiben mit $X_1 \sim \nu_\vartheta$, und ν_ϑ ist Mitglied der Familie $(\nu_\vartheta)_{\vartheta\in\Theta}$ der quadratintegrierbaren Verteilungen mit positiver Varianz. In diesem nichtparametrischen Modell sind wir an einem Konfidenzintervall für $\mu_\vartheta := \mathbb{E}_\vartheta[X_1]$ interessiert.

Die Grundidee ist die Folgende: In Beispiel 6.2 haben wir ausgenutzt, dass der Mittelwert unabhängiger und normalverteilter Zufallsvariablen auch normalverteilt ist. Das gilt für allgemeine Verteilungen nicht. Allerdings ist der Mittelwert nach dem Zentralen Grenzwertsatz (Satz 2.11) *asymptotisch* normalverteilt. Das ist der Schlüssel und erklärt, warum das Konfidenzintervall im nichtparametrischen Fall die gleiche Struktur hat wie im Rahmen der Normalverteilungsannahme.

Beispiel 6.3 (Asymptotisches Konfidenzintervall für den Erwartungswert bei bekannter Varianz) *Seien $X_1, X_2 \dots$ unabhängige und identisch verteilte Zufallsvariable mit $X_1 \sim \nu_\vartheta$, und ν_ϑ ist Mitglied der Familie $(\nu_\vartheta)_{\vartheta\in\Theta}$ der quadratintegrierbaren Verteilungen mit fester, bekannter (!) Varianz $\mathbb{V}\mathrm{ar}_\vartheta(X_1) = \sigma^2 > 0$.*

Weiter sei q das $(1-\alpha/2)$-Quantil der Standardnormalverteilung. Dann ist

$$I_n(\mathbf{x}_n) := \left[\bar{x}_n - q\cdot\frac{\sigma}{\sqrt{n}}, \bar{x}_n + q\cdot\frac{\sigma}{\sqrt{n}}\right] \tag{6.4}$$

ein asymptotisches $(1-\alpha)$-Konfidenzintervall für $\mu_\vartheta := \mathbb{E}_\vartheta[X_1]$, genauer gilt $\mathbb{P}_\vartheta(I_n(\mathfrak{X}_n) \ni \mu_\vartheta) \to 1-\alpha$, für $n\to\infty$. Denn aufgrund des Zentralen Grenzwertsatzes gilt $(\bar{X}_n - \mu_\vartheta)/(\sigma/\sqrt{n}) \xrightarrow{d_\vartheta} N(0,1)$ für $n\to\infty$. Dabei bedeutet die Indizierung mit ϑ in der Schreibweise ‚$\xrightarrow{d_\vartheta}$' die Konvergenz in Verteilung unter der Annahme, dass ν_ϑ die wahre zugrunde liegende Verteilung der Zufallsvariablen ist. Damit haben wir den Mittelwert $\bar{X}_n$ durch μ_ϑ korrekt zentriert. Es folgt für $n\to\infty$

$$\mathbb{P}_\vartheta(I_n(\mathfrak{X}_n) \ni \mu_\vartheta) = \stackrel{(*)}{\cdots} = \mathbb{P}_\vartheta\left(\frac{\bar{X}_n - \mu_\vartheta}{\sigma/\sqrt{n}} \in [-q, q]\right) \longrightarrow 1-\alpha,$$

wobei in $(*)$ die gleichen Umformungen wie in Beispiel 6.2 durchgeführt wurden.

Schätzung der Standardabweichung und der Begriff des Standardfehlers Dank des Zentralen Grenzwertsatzes sind wir zur Konstruktion eines Konfidenzintervalls für den Erwartungswert nun nicht mehr auf die Normalverteilungsannahme angewiesen – zumindest wenn die Anzahl der Beobachtungen n ‚hinreichend groß' ist. Allerdings ist in der Praxis die Varianz der Beobachtungen praktisch immer unbekannt. Ziel ist daher, auch bei unbekannter Varianz ein approximatives Konfidenzintervall zu konstruieren, vgl. (6.3). Die Grundidee ist, die Größe σ durch einen konsistenten Schätzer zu ersetzen. Dass wir das dürfen, ist eine Konsequenz des Satzes von Slutsky (Satz 2.12).

Beispiel 6.4 (Asymptotisches Konfidenzintervall für den Erwartungswert)
Seien $X_1, X_2 \dots$ unabhängige und identisch verteilte Zufallsvariable mit $X_1 \sim \nu_\vartheta$, und ν_ϑ ist Mitglied der Familie $(\nu_\vartheta)_{\vartheta\in\Theta}$ der quadratintegrierbaren Verteilungen mit positiver Varianz. Zudem sei $(S_n)_{n=1,2,\dots}$ eine konsistente Folge von Schätzern für $\sigma_\vartheta := (\mathbb{V}ar_\vartheta(X_1))^{1/2}$ und q das $(1-\alpha/2)$-Quantil der Standardnormalverteilung. Dann ist

$$I_n(\mathbf{x}_n) := \left[\bar{x}_n - q\cdot\frac{S_n(\mathbf{x}_n)}{\sqrt{n}}, \bar{x}_n + q\cdot\frac{S_n(\mathbf{x}_n)}{\sqrt{n}}\right] \tag{6.5}$$

ein asymptotisches $(1-\alpha)$-Konfidenzintervall für $\mu_\vartheta := \mathbb{E}_\vartheta[X_1]$, genauer ist $\mathbb{P}_\vartheta(I_n(\mathfrak{X}_n) \ni \mu_\vartheta) \to 1-\alpha$, für $n\to\infty$.

Denn aufgrund des Satzes von Slutsky bleibt die Konvergenzaussage im Zentralen Grenzwertsatz erhalten, wenn man die Standardabweichung durch eine Folge von Zufallsvariablen ersetzt, die in Wahrscheinlichkeit gegen ebendiese strebt, genauer gilt für alle $\vartheta\in\Theta$

$$\frac{\bar{X}_n - \mu_\vartheta}{S_n(\mathfrak{X}_n)/\sqrt{n}} = \frac{\sigma_\vartheta}{S_n(\mathfrak{X}_n)}\,\frac{\bar{X}_n-\mu_\vartheta}{\sigma_\vartheta/\sqrt{n}} \xrightarrow{d_\vartheta} N(0,1), \tag{6.6}$$

für $n\to\infty$, denn $\sigma_\vartheta/S_n(\mathfrak{X}_n) \xrightarrow{\mathbb{P}_\vartheta} 1$. Wieder finden wir für $n\to\infty$

$$\mathbb{P}_\vartheta(I_n(\mathfrak{X}_n)\ni\mu_\vartheta) \overset{(*)}{=} \cdots = \mathbb{P}_\vartheta\left(\frac{\bar{X}_n-\mu_\vartheta}{S_n(\mathfrak{X}_n)/\sqrt{n}} \in [-q,q]\right) \longrightarrow 1-\alpha,$$

wobei in $(*)$ wieder die Umformungen aus Beispiel 6.2 durchzuführen sind.

Die Größe $S_n(\mathbf{x}_n)/\sqrt{n}$ ist ein Schätzer für die Schwankung des Mittelwertes. Man nennt ihn auch einen *Standardfehler des Mittelwertes*. In vorheriger Formulierung des Konfidenzintervalles kann ein beliebiger konsistenter Schätzer S_n für die Standardabweichung σ_ϑ eingesetzt werden. Wir erinnern an die prominente empirische Standardabweichung s_n, vgl. (3.3), sowie Beispiel 5.4i. Die Größe

$$sem_n(\mathbf{x}_n) := \frac{s_n(\mathbf{x}_n)}{\sqrt{n}}$$

wird häufig sogar als *der* Standardfehler des Mittelwertes bezeichnet. Wir betonen, dass wir den sem_n als Schätzer für die Standardabweichung $\sigma_\vartheta/\sqrt{n}$ des Mittelwertes verstehen und dass der Mittelwert eine Statistik ist – es geht hier *nicht* um die Schwankung σ_ϑ der Individualbeobachtungen. In folgender Definition verallgemeinern wir den Begriff des Standardfehlers des Mittelwertes auf beliebige quadratintegrierbare Statistiken.

Definition 6.5 (Standardfehler)
Es sei ein statistisches Modell gegeben durch einen Zufallsvektor $\mathfrak{X}_\infty = (X_1, X_2, \ldots)^t$ und eine Verteilungsfamilie $(\nu_\vartheta)_{\vartheta\in\Theta}$. Für $n = 1, 2\ldots$ sei im Modell der Restriktion auf die ersten n Komponenten ($\mathfrak{X}_n = (X_1, \ldots, X_n)^t$)

1. *T_n eine Statistik mit existierender positiver Varianz $\mathbb{V}ar_\vartheta(T_n(\mathfrak{X}_n)) > 0$,*
2. *se_n ein Schätzer der Standardabweichung $(\mathbb{V}ar_\vartheta(T_n(\mathfrak{X}_n)))^{1/2}$ der Statistik T_n.*

Dann heißt $(se_n)_{n=1,2\ldots}$ ein Standardfehler von $(T_n)_{n=1,2,\ldots}$ (engl. standard error), falls für alle $\vartheta \in \Theta$ gilt

$$\frac{se_n(\mathfrak{X}_n)}{(\mathbb{V}ar_\vartheta(T_n(\mathfrak{X}_n)))^{1/2}} \xrightarrow{\mathbb{P}_\vartheta} 1 \quad f\ddot{u}r \quad n \to \infty. \tag{6.7}$$

Wir sagen auch kurz, dass se_n ein Standardfehler von T_n ist. Wir erinnern, dass ein Schätzer im Grunde eine beliebige Abbildung ist. Die Eigenheit des Standardfehlers $se_n(\mathfrak{X}_n)$ ist es nun, die zu schätzende Kenngröße – also die Standardabweichung des Schätzers $T_n(\mathfrak{X}_n)$ – für wachsende Anzahl n von Beobachtungen im Sinne von (6.7) gutartig zu schätzen. Dass diese definierende Eigenschaft Sinn ergibt, haben wir schon oben in (6.6) bei der Konstruktion des asymptotischen Konfidenzintervalls gesehen, denn sie erlaubt das Ersetzen der unbekannten Standardabweichung $\sigma_\vartheta/\sqrt{n}$ des Mittelwerts mit einem Standardfehler, etwa $sem_n(\mathfrak{X}_n)$.

Schließlich merken wir an, dass der Standardfehler des Mittelwertes sem_n oder auch allgemeiner ein beliebiger Standardfehler des Mittelwertes $S_n/\sqrt{n}$ die definierende Eigenschaft eines Standardfehlers (6.7) erfüllt. Denn aufgrund der Konsistenz von S_n für σ_ϑ folgt für $n \to \infty$

$$\frac{S_n(\mathfrak{X}_n)/\sqrt{n}}{(\mathbb{V}ar_\vartheta(\bar{X}_n))^{1/2}} = \frac{S_n(\mathfrak{X}_n)}{\sigma_\vartheta} \xrightarrow{\mathbb{P}_\vartheta} 1.$$

Wir formulieren die Hauptaussage dieses Abschnittes erneut: Es gilt im Rahmen des Modells aus Beispiel 6.4, dass

$$I_n(\mathbf{x}_n) := [\bar{x}_n - q \cdot sem(\mathbf{x}_n), \bar{x}_n + q \cdot sem(\mathbf{x}_n)]$$

ein asymptotisches $(1-\alpha)$-Konfidenzintervall für den Erwartungswert der Individualbeobachtung ist, wobei q das $(1-\alpha/2)$-Quantil der Standardnormalverteilung bezeichnet.

6.4 Dialog: Das Konfidenzintervall

Ein guter Freund besucht Sie und möchte Ihren Rat als Statistiker. Er ist Biologe und war mit seiner Arbeitsgruppe in den Semesterferien zur Feldforschung in Chile. Dort hat er im Rahmen seiner Abschlussarbeit Insekten einer bestimmten Spezies untersucht. Er hat – in der Zeit, in der er nicht gerade im Meer schwimmen oder in den Anden wandern war – tatsächlich $n = 100$ Insekten beobachtet und u. a. deren Gewicht $\mathbf{x} = (3.65, 5.14, 4.11, 4.42, \ldots, 5.23)^t$ in Gramm bestimmt. Das sind die Beobachtungen aus Abb. 3.1. Während Sie zusammen eine Tasse des exzellenten, von Ihrem Kollegen importierten Kaffees trinken, spielt sich zwischen Ihnen (**S**tatistiker) und Ihrem Freund (**B**iologe) folgender Dialog ab:

B: ‚Jetzt habe ich all diese Beobachtungen gemacht, aber ich weiß nicht wirklich, was ich mit ihnen anfangen soll. Echt keine Ahnung. . .! Ich würde eigentlich gern die ‚beobachteten Gewichte der Insekten‘ statistisch auswerten. Dazu hab ich auch schon ein bisschen was mit Excel gerechnet. . . So richtig weiß ich aber eigentlich nicht, was ich da tue, wenn ich ehrlich bin. Kannst du dir meine Berechnungen mal anschauen?‘

Da Sie zunächst erst einmal sehen wollen, womit Sie es zu tun haben, schmeißen Sie Ihren Rechner an und erstellen Abb. 3.1a oder b. Sie machen Ihren Freund auf diesen grundlegenden Aspekt der Datenanalyse aufmerksam:

S: Bevor wir anfangen zu rechnen, sollten wir die Beobachtungen erstmal grafisch darstellen!

Ihr Freund ist von den Abbildungen zwar durchaus angetan, weiß aber nicht recht, wie er sie interpretieren soll.

B: Okay, alles klar, das sind ja schöne Punkte. . . Aber was sagt mir das?

Sie erklären, dass die Beobachtungen sich näherungsweise glockenförmig verteilen, und erzählen Ihrem Kumpan die Story rund um die Waage und die 2/3-Regel. Dann schließen Sie:

S: . . . daher können wir sogar direkt den Mittelwert als etwa 4.5 und die Standardabweichung als etwa 1 ablesen.

B: Wow, und das ganz ohne zu rechnen. . .

Ihr Freund denkt eine Weile nach und meldet dann Bedenken an.

B: Mein Arbeitsgruppenleiter forscht schon etwas länger zu diesen Insekten. Er behauptet, das mittlere Gewicht in der Population sei 5.5 g.

Sie erinnern sich an das Konfidenzintervall für den Erwartungswert (6.3). Bei $n = 100$ Beobachtungen und einer geschätzen Standardabweichung von 1 ist der Standardfehler 1/10 und das auf den Beobachtungen basierende 95 %-Konfidenzintervall etwa $I(\mathbf{x}) \approx [4.3, 4.7]$. Auf Basis dieses Wissens erläutern Sie:
S: Das berechnete 95 %-Konfidenzintervall überdeckt den behaupteten Wert 5.5 *nicht*.
Ihr Freund schließt daraufhin etwas übereilig:
B: Ach stimmt, ich erinnere mich an diese Intervalle. Das Ergebnis bedeutet also, dass mein Chef recht hat!
Sie korrigieren ihn.
S: Nein!
B: Oh, dann heißt es, dass mein Chef unrecht hat.
S: Auch das können wir nicht sagen! Leider können wir in der Statistik viel weniger gewichtige Aussagen machen, als die meisten Leute denken.
B: Stimmt ja, ihr rechnet ja immer mit Wahrscheinlichkeiten. Dass das Intervall den behaupteten Parameter nicht überdeckt, bedeutet dann also, dass der wahre Populationsmittelwert wahrscheinlich – ähhh, mit einer Wahrscheinlichkeit von 95 % – in dem Intervall liegt, oder?
Wieder müssen Sie ihn korrigieren.
S: Ich muss dich leider enttäuschen. Wir modellieren Größen wie das Populationmittel nicht als zufällig. Und daher können wir solchen Parametern keine Wahrscheinlichkeit zuordnen.
(Sie haben vielleicht schon mal von Bayes'scher Statistik gehört. Dort werden auch die Parameter als zufällig modelliert. Aber das ist eine ganz andere Welt, und damit wollen Sie Ihren Freund jetzt nicht verwirren.)
Ihr Freund runzelt die Stirn. Nun möchte er es genau wissen.
B: Aber was bedeutet es denn nun, dass das Intervall den behaupteten Wert nicht überdeckt?
S: Das bedeutet lediglich, dass etwas Unwahrscheinliches beobachtet wurde, falls dein Chef recht hat. Genauer: Unter der Annahme, dass er recht hat, ist etwas eingetreten, das nur in 5 % der Fälle eintritt. Das empfinden wir als unwahrscheinlich. Es sagt aber nichts darüber aus, ob er tatsächlich oder auch nur wahrscheinlich recht hat oder nicht.
B: Kannst du das noch etwas erläutern? Für einen Laien wie mich?
S: Ich kann's versuchen... Also, wir machen eigentlich Modellannahmen, bei denen wir davon ausgehen, dass deine Beobachtungen Produkte des Zufalls sind. Ein zufälliges Konfidenzintervall ist nun aus eben jenem Zufall konstruiert. Es ist gerade so gebaut, dass es das wahre Populationsmittel – das nennen wir im Rahmen des Modells auch Erwartungswert – mit 95 %-iger Wahrscheinlichkeit überdeckt. Also, wenn dein Chef recht hat und 5.5 der wahre Populationsmittelwert ist, dann wird die 5.5 auch mit einer Wahrscheinlichkeit von 95 % überdeckt. Wir interpretieren nun deine Beobachtungen als Realisierungen des Zufalls, d. h., der Zufall ist schon eingetreten, und folglich sind die Beobachtungen und damit auch das Intervall [4.3, 4.7] feste Größen. Da das Intervall den

Wert 5.5 nicht überdeckt, sind deine Beobachtungen als ein unwahrscheinliches Ergebnis des Zufalls zu werten – eines der „extremen“ Ereignisse, welche nur in 5 % der Fälle auftreten.
B: Und was bedeutet das?
S: Leider nicht viel. Es bedeutet eben nur, dass die Beobachtungen nicht mit der Behauptung Deines Chefs verträglich sind.
B: Und zu diesem Schluss kommen wir aufgrund der diskutierten Unwahrscheinlichkeit des beobachteten Intervalls unter der Annahme, dass mein Chef recht hat?
S: Ja, genau!
B: Cool! Und umgekehrt, wenn jetzt jemand ankäme und behauptete, das wahre Populationsmittel läge bei 4.6, bedeutete das dann im Gegenzug ebenfalls nicht, dass er recht hat, sondern nur, dass etwas Wahrscheinliches eingetreten ist?
S: Genau! Allerdings rede ich lieber davon, dass, wenn die Behauptung stimmt, nichts Unwahrscheinliches beobachtet wurde, wir also aufgrund der Beobachtungen keinen Anlass haben, an der Behauptung zu zweifeln. In diesem Sinne sind die Beobachtungen mit dem behaupteten Mittel von 4.6 verträglich. Wir nutzen diese Formulierung der UNwahrscheinlichkeit, weil wir davon ausgehen, dass diese Behauptungen sowieso immer falsch sind. Du glaubst wohl selber nicht daran, dass das wahre Populationsmittel exakt 4.6 ist. So gesehen ist eigentlich immer nur die Frage, ob uns die Beobachtungen *genug* Anlass geben, an der Behauptung zu zweifeln, oder ob sie es nicht tun.
Ihr Freund wirkt alles in allem sehr zufrieden. Aber etwas wundert ihn.
B: Eine Frage noch. Was soll die ganze Sache mit Wörtern wie „etwa“ oder „ungefähr“ und all diese dubiosen Schätzungen aus Bildern? Du bist doch eigentlich auch MathematikerIn. Wollt ihr es nicht immer ganz genau wissen?
S: Wir müssen entscheiden, an welcher Stelle es sinnvoll ist, mathematische Genauigkeit an den Tag zu legen. Die statistische Theorie sollten wir unbedingt mathematisch exakt formulieren können. In der Anwendung ist das anders.
Um das Ganze ein bisschen anschaulicher zu machen, bringen Sie noch ein Beispiel.
S: Denke etwa mal darüber nach, inwieweit sich obiges Resultat änderte, wenn wir nicht 4.5 als Mittelwert aus der Grafik erhielten, sondern einen viel genaueren Wert, nämlich 4.506372, aus den Beobachtungen errechneten. Und wenn wir nicht die Standardabweichung aus der Grafik als 1 schätzten, sondern als 0.9738418 berechneten.
Ihr Freund versteht sofort, was Sie meinen.
B: Das würde am Konfidenzintervall ja nur minimal etwas ändern, und das Intervall wäre immer noch weit vom behaupteten Wert 5.5 entfernt.
S: Richtig. Das Intervall wäre immer noch etwa acht Standardfehler von 5.5 entfernt. Das ist weit, wenn man bedenkt, dass die typische Abweichung des Mittelwerts etwa ein Standardfehler ist. Schließlich bliebe die Interpretation die gleiche: Falls die Behauptung stimmt, dann ist etwas Unwahrscheinliches eingetreten.
Sie fügen hinzu:

S: Denk' in diesem Zusammenhang auch darüber nach, dass es an uns liegt zu entscheiden, was „unwahrscheinlich" bedeutet. Wir könnten auch ein 99 %-Konfidenzintervall basteln. Das ist dann größer und so konstruiert, dass es den wahren Parameter mit 99 %-iger Wahrscheinlichkeit überdeckt.
Ihr Freund ist von so viel Gedankenakrobatik sichtlich erschöpft. Trotzdem scheint er froh, ein bisschen was verstanden zu haben. Er seufzt.
B: Puhh, gar nicht so einfach. Aber ich glaube, ich hab' es jetzt einigermaßen kapiert: Im Grunde möchte ich die Unwahrscheinlichkeit meiner Beobachtungen beurteilen. Jetzt habe ich aber genug davon. Noch ein Käffchen?
Das ist ganz in Ihrem Sinne. Erfreut stimmen Sie zu.

6.5 Ein Konfidenzintervall für den Median

Während Sie beide Ihren Kaffee trinken, fällt Ihrem Freund noch etwas auf.

B: Super, dass ich das jetzt verstanden habe, dann kann ich so ein 95 %-Konfidenzintervall ja auch für alle anderen Variablen berechnen, oder? Mein Chef interessiert sich nämlich auch für die Entfernung der Tiere zu ihrem Bau. Die habe ich auch gemessen.
Leider müssen Sie seine Euphorie an dieser Stelle ein wenig ausbremsen.
S: Im Prinzip ist das eine gute Idee, aber man muss leider aufpassen und jedes Mal alle Schritte der Reihe nach machen. Beim Gewicht haben wir ja auch die Beobachtungen zuerst dargestellt... Das ist ein wichtiger Schritt.
B: Und warum? Die Rechnung kann ich doch eigentlich immer machen, und ich habe mal gelesen, dass der Mittelwert immer normalverteilt ist, was kann denn dann noch schiefgehen?
Um Ihrem Freund das Ganze zu veranschaulichen, erstellen Sie schnell eine weitere Grafik. Sie klappen Ihren Rechner auf, suchen die Variable Entfernung und erstellen aus ihr zum Beispiel Abb. 3.2.
S: Schau mal, diese Beobachtungen sind zum Beispiel nicht symmetrisch verteilt: Die meisten Tiere haben sich nah am Bau aufgehalten, nur wenige in größerem Abstand. Die Verteilung ist also ziemlich asymmetrisch, der Mittelwert ist viel größer als die meisten Beobachtungen. Er liefert daher keine gute Beschreibung für die Lage deiner Beobachtungen. Aber wir könnten stattdessen den Median betrachten.

Wir verlassen hier das Gespräch, um ein Konfidenzintervall für den Median theoretisch herzuleiten. Wir betrachten wieder ein exaktes und ein asymptotisches Konfidenzintervall. Letzteres beruht auf der asymptotischen Verteilung des Medians. Bei dem exakten Konfidenzintervall ist besonders interessant, dass man es im Gegensatz zum Mittelwert für

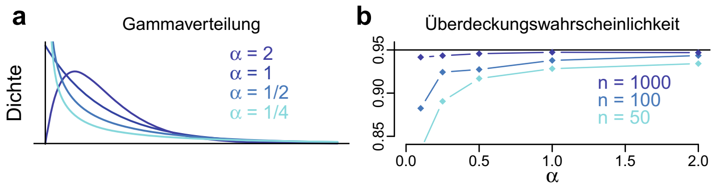


Abb. 6.3 **a** Dichten der Gammaverteilung mit Erwartungswert 1 und verschiedenen Formparametern α. **b** Überdeckungswahrscheinlichkeit des asymptotischen 95 %-Konfidenzintervalls für den Erwartungswert im Falle unabhängiger und identisch Gamma-verteilter Zufallsvariablen $X_1, \ldots, X_n$ mit $X_1 \sim \gamma(\alpha, \lambda)$ bei Erwartungswert 1 und Formparameter α. Für festes n fällt die Überdeckungswahrscheinlichkeit mit fallendem α

beliebige Verteilungen ohne Kenntnis weiterer Parameter erstellen kann. Man kann also für den Median ein Konfidenzintervall konstruieren, das die Überdeckungswahrscheinlichkeit mindestens $1-\alpha$ für festes n einhält. Beim Erwartungswert sind wir dagegen auf das asymptotische Konfidenzintervall angewiesen, dessen Überdeckungswahrscheinlichkeit von n abhängt, und für festes n von der Form der Verteilung (siehe Abb. 6.3).

Konstruktion: Ein Konfidenzintervall für den Median Es seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariable mit $X_1 \sim \nu_\vartheta$, und ν_ϑ ist Mitglied der Familie $(\nu_\vartheta)_{\vartheta \in \Theta}$ der Verteilungen mit stetiger Verteilungsfunktion. Weiter sei m_ϑ ein Median der Verteilung ν_ϑ und $\alpha \in (0, 1)$. Wir suchen ein möglichst kleines Intervall $I(\mathbf{x})$, sodass für alle $\vartheta \in \Theta$ gilt

$$\mathbb{P}_\vartheta(I(\mathfrak{X}) \ni m_\vartheta) \overset{(*)}{\geq} 1 - \alpha.$$

Wir betrachten dazu Intervalle, die aus den Ordnungsstatistiken gebildet werden. Sei $\mathbf{x} = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ und $(x_{(1)}, \ldots, x_{(n)})^t$ die Ordnungsstatistik von $\mathbf{x}$. Dann setzen wir

$$I^{(k)}(\mathbf{x}) := \left[x_{(k)}, x_{(n-k+1)}\right]$$

für $k = 1, \ldots, \lfloor n/2 \rfloor$, d. h., wir betrachten das Intervall von der k-kleinsten bis zur $(n-k+1)$-größten Beobachtung.

Ziel ist es, ein möglichst kleines Intervall $I^{(k)}$ zu konstruieren, sodass die Überdeckungseigenschaft $(*)$ gerade noch gilt. Der Trick ist, dass wir über die definierende Eigenschaft des Medians die Binomialverteilung ins Spiel bringen: Jede Beobachtung X_i liegt mit Wahrscheinlichkeit 1/2 links und mit Wahrscheinlichkeit 1/2 rechts des Medians. Hier nutzen wir die Stetigkeit der Verteilungsfunktion, bei der es keine Punktmassen gibt, insbesondere am Median nicht. Es sei $Y \sim b(n, 1/2)$ und unabhängig von $\mathfrak{X}$. Für $k = 1$ gilt dann

$$\mathbb{P}_\nu\left(I^{(1)}(\mathfrak{X}) \not\ni m_\nu\right) = \mathbb{P}_\nu(X_{(1)} > m_\nu) + \mathbb{P}_\nu(X_{(n)} < m_\nu) = 2\left(\frac{1}{2}\right)^n = 2\mathbb{P}(Y \leq 0),$$

denn $I^{(1)}$ überdeckt den Median m_ϑ nicht, wenn entweder alle Beobachtungen größer als m_ϑ oder alle Beobachtungen kleiner als m_ϑ sind. Diese Ereignisse sind disjunkt – offenbar kann nicht beides gleichzeitig eintreten –, und jedes tritt mit Wahrscheinlichkeit $(1/2)^n$ ein. Für $k = 2$ argumentieren wir

$$\begin{aligned}\mathbb{P}_\vartheta\left(I^{(2)}(\mathfrak{X}) \not\ni m_\vartheta\right) &= \mathbb{P}_\vartheta(\{\text{höchstens ein } X_i \text{ ist kleiner als } m_\vartheta\}) \\ &\quad + \mathbb{P}_\vartheta(\{\text{höchstens ein } X_i \text{ ist größer als } m_\vartheta\}) \\ &= 2 \cdot \left[\binom{n}{0}\left(\frac{1}{2}\right)^n + \binom{n}{1}\left(\frac{1}{2}\right)^n\right] \\ &= 2 \cdot \mathbb{P}(Y \leq 1).\end{aligned}$$

Wir erkennen die ‚Binomialstruktur'. Für allgemeines k gilt

$$\mathbb{P}_\vartheta\left(I^{(k)}(\mathfrak{X}) \not\ni m_\nu\right) = 2 \cdot \mathbb{P}(Y \leq k-1). \tag{6.8}$$

Jetzt suchen wir ein möglichst kleines Intervall $I^{(k)}$ ($\leftrightarrow$ möglichst großes k), sodass der Ausdruck (6.8) gerade noch kleiner α bleibt. Wir suchen also $k^* = max\{k|\mathbb{P}(Y \leq k-1) < \alpha/2\}$ und setzen $I := I^{(k^*)}$.

Wir bemerken nochmal, dass wir kaum Annahmen an die Verteilung gemacht haben. Eigentlich haben wir die Stetigkeit der Verteilungsfunktion nur genutzt, um Punktmassen am Median auszuschließen. Wir könnten also auch diese schwächere Bedingung fordern. Zudem kann man über analoge Argumente ein Konfidenzintervall für ein p-Quantil konstruieren.

Asymptotische Normalität des Medians In Abschn. 6.3 haben wir ein asymptotisches Konfidenzintervall für den Erwartungswert konstruiert. Dieses basierte auf der asymptotischen Normalität des Mittelwertes. Der entsprechende Baustein zur Konstruktion eines asymptotischen Konfidenzintervalls für den Median der Verteilung ist die asymptotische Normalität des Stichprobenmedians.

Satz 6.6 (Asymptotische Normalität des Medians)
Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariable mit Verteilungsfunktion F, und sei m ein Median der assoziierten Verteilung. Für $n = 1, 2, \ldots$ sei $\mathfrak{X}_n = (X_1, \ldots, X_n)^t$ und $M_n(\mathfrak{X}_n)$ der Stichprobenmedian der ersten n Zufallsvariablen. Ist F am Median m differenzierbar und gilt $F^{'}(m) > 0$, dann gilt für $n \to \infty$

$$\sqrt{n}(M_n(\mathfrak{X}_n) - m) \xrightarrow{d} N\left(0, \frac{1}{(2F^{'}(m))^2}\right).$$

Wir bemerken, dass die Bedingung $F'(m) > 0$ die Eindeutigkeit des Medians impliziert.

Beweisidee Wir stellen $M_n(\mathfrak{X}_n)$ als Summe unabhängiger und identisch verteilter Zufallsvariablen dar und wenden auf diese den Zentralen Grenzwertsatz an. Etwas genauer: Sei $Z \sim N(0,1)$. Nach Definition der Verteilungskonvergenz ist zu zeigen, dass $\mathbb{P}(\sqrt{n}(M_n(\mathfrak{X}_n) - m) \le z) \to \mathbb{P}(Z \le 2zF'(m))$ für alle $z \in \mathbb{R}$. Für n ungerade gilt

$$
\begin{aligned}
\{\sqrt{n}(M_n(\mathfrak{X}_n) - m) \le z\} &= \left\{X_{((n+1)/2)} \le m + \frac{z}{\sqrt{n}}\right\} \\
&= \left\{\text{mindestens } \frac{n+1}{2} \text{ der } X_i \text{ sind } \le m + \frac{z}{\sqrt{n}}\right\} \\
&= \left\{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le m + z/\sqrt{n}\}} \ge \frac{n+1}{2}\right\} \\
&= \left\{\frac{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le m + z/\sqrt{n}\}} - np_n}{\sqrt{np_n(1-p_n)}} \ge \frac{(n+1)/2 - np_n}{\sqrt{np_n(1-p_n)}}\right\}, \qquad (6.9)
\end{aligned}
$$

wobei wir $p_n := F(m + z/\sqrt{n})$ gesetzt haben. Die Indikatorvariablen sind für festes n unabhängig und $ber(p_n)$-verteilt, sodass wir wieder ein Binomialargument nutzen. So kommt der Zentrale Grenzwertsatz ins Spiel. Für $n \to \infty$ gilt

$$
Z_n := \frac{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le m + z/\sqrt{n}\}} - np_n}{\sqrt{np_n(1-p_n)}} \xrightarrow{d} N(0,1).
$$

Da die Summanden unter verschiedenen n zwar unabhängig, aber nicht identisch verteilt sind, braucht man hier eine Verallgemeinerung des Zentralen Grenzwertsatzes von Lindeberg, siehe zum Beispiel Feller (1971). Für die asymptotische Varianz errechnen wir für $n \to \infty$

$$
\begin{aligned}
c_n := \frac{(n+1)/2 - np_n}{\sqrt{np_n(1-p_n)}} &= \frac{1}{\sqrt{p_n(1-p_n)}} \frac{-(p_n - 1/2)}{1/\sqrt{n}} + \frac{1/2}{\sqrt{np_n(1-p_n)}} \\
&= \frac{1}{\sqrt{p_n(1-p_n)}} \frac{-(F(m+z/\sqrt{n}) - F(m))}{1/\sqrt{n}} + \frac{1/2}{\sqrt{np_n(1-p_n)}} \\
&= \underbrace{\frac{-z}{\sqrt{p_n(1-p_n)}}}_{\to -2z} \underbrace{\frac{F(m+z/\sqrt{n}) - F(m)}{z/\sqrt{n}}}_{\to F'(m)} + \underbrace{\frac{1/2}{\sqrt{np_n(1-p_n)}}}_{\to 0} \\
&\longrightarrow -2zF'(m).
\end{aligned}
$$

Hier haben wir insbesondere ausgenutzt, dass $F(m) = 1/2$ gilt und F am Median differenzierbar ist. Mit dem Satz von Slutsky 2.12 folgt für $n \to \infty$, dass $Z_n^* := Z_n - c_n - 2zF'(m) \xrightarrow{d} Z$ und damit

$$\mathbb{P}(\sqrt{n}(M_n(\mathfrak{X}_n) - m) \leq z) \overset{(6.9)}{=} \mathbb{P}(Z_n \geq c_n) = \mathbb{P}(Z_n^* \geq -2zF'(m)) \xrightarrow{d} \mathbb{P}(Z \geq -2zF'(m)),$$

wobei der Grenzwert aus Symmetriegründen $\mathbb{P}(Z \leq 2zF'(m))$ entspricht. Für gerades n gilt $X_{(n/2+1)} \geq M_n(\mathfrak{X}_n) \geq X_{(n/2)}$, sodass

$$\mathbb{P}(\sqrt{n}(X_{(n/2+1)} - m) \leq z) \leq \mathbb{P}(\sqrt{n}(M(\mathfrak{X}) - m) \leq z) \leq \mathbb{P}(\sqrt{n}(X_{(n/2)} - m) \leq z).$$

Nun lässt sich aber analog zum Fall der ungeraden n die Konvergenz der linken und rechten Seite gegen $\mathbb{P}(Z \leq 2zF'(m))$ zeigen.

Aus Satz 6.6 lassen sich noch einige schöne Eigenschaften des Medians ableiten.

Die Varianz von M_n Laut Satz 6.6 ist die asymptotische Varianz von $M_n(\mathfrak{X}_n)$ gegeben durch $1/(4nF'(m)^2)$, bzw. $1/(4nf(m)^2)$, falls X_1 Dichte f besitzt. Interessanterweise wird also die Variabilität des Stichprobenmedians von einer völlig anderen Eigenschaft der Verteilung – nämlich von der Dichte am Median – bestimmt als die Variabilität des Mittelwerts, die stattdessen von σ abhängt. Abb. 6.4 illustriert dies an verschiedenen Verteilungsbeispielen. So kann in manchen Fällen (Abb. 6.4a) der Stichprobenmedian stark schwanken, während der Mittelwert eine geringe Variabilität zeigt. In anderen Fällen (Abb. 6.4d) kann es umgekehrt sein. Im Fall der Normalverteilung steigt die Variabilität beider Größen mit σ (Abb. 6.4b und c).

Konsistenz von M_n Wir bemerken abschliessend, dass wir unter den Bedingungen des Satzes 6.6 auch sofort die Konsistenz von M_n für m folgern können. Denn mit dem Satz von Slutsky (Satz 2.12) gilt für $n \to \infty$

$$M_n(\mathfrak{X}_n) - m = \frac{1}{\sqrt{n}} \sqrt{n}(M_n(\mathfrak{X}_n) - m) \xrightarrow{d} 0.$$

Da aber die Konvergenz in Verteilung gegen eine Konstante auch die Konvergenz in Wahrscheinlichkeit impliziert (siehe Lemma 2.14), folgt die Konsistenz von $M_n(\mathfrak{X}_n)$ für m.

In der Tat gilt allgemeiner für den Stichprobenmedian $M_n(\mathfrak{X}_n)$ von unabhängigen und identisch verteilten Zufallsvariablen mit Verteilungsfunktion F sogar, dass

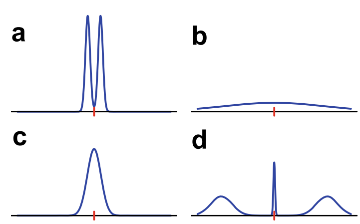


Abb. 6.4 Variabilität von Mittelwert und Median. Die asymptotische Varianz des Mittelwerts ist klein bei kleiner Standardabweichung (**a, c**), während die asymptotische Varianz des Medians klein ist, wenn die Dichte am Median groß ist (**c, d**)

$$\liminf_{n\to\infty} M_n(\mathfrak{X}_n) \geq q_{1/2}^- \qquad \text{sowie} \qquad \limsup_{n\to\infty} M_n(\mathfrak{X}_n) \leq q_{1/2}^+$$

mit Wahrscheinlichkeit 1. Dabei bezeichnet $Q_{1/2} = [q_{1/2}^-, q_{1/2}^+]$ das Intervall der Mediane der mit F assoziierten Verteilung. Das impliziert die starke Konsistenz von $M_n(\mathfrak{X}_n)$ für m, sofern m der eindeutige Median der Verteilung ist. Die Notwendigkeit der Eindeutigkeit erscheint plausibel, denn andernfalls wäre m Element des Intervalls $Q_{1/2}$, in welchem gar keine Ereignisse eintreten. Der Stichprobenmedian $M_n(\mathfrak{X}_n)$ hätte gar nicht die Möglichkeit sich m anzunähern, denn zumindest für ungerades n läge $M_n(\mathfrak{X}_n)$ immer außerhalb dieses Intervalls.

6.6 Ein Konfidenzband für die Verteilungsfunktion

Zusätzlich zum exakten oder approximativen Konfidenzintervall für einen Parameter kann man auch für eine ganze Funktion einen Konfidenzbereich definieren. Im Folgenden seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariable mit $X_1 \sim \nu_\vartheta$, und ν_ϑ ist Mitglied der Familie $(\nu_\vartheta)_{\vartheta\in\Theta}$ aller reellwertigen Verteilungen. Bezüglich einer Verteilung ν_ϑ betrachten wir nun die assoziierte Verteilungsfunktion F_ϑ als die unbekannte abgeleitete Kenngröße, $F_\vartheta(z) = \mathbb{P}_\vartheta(X_1 \leq z)$ für alle $z \in \mathbb{R}$. Bislang haben wir einen reellwertigen Parameter wie etwa den Erwartungswert oder den Median betrachtet und dafür ein Konfidenz*intervall* konstruiert. Hier sind wir an einer ganzen Funktion F_ϑ interessiert und suchen anschaulich ein aus den Beobachtungen konstruiertes *Band,* das F_ϑ mit großer Wahrscheinlichkeit überdeckt, siehe Abb. 6.5. Es mag nicht überraschen, dass in die Konstruktion dieses Objekts die empirische Verteilungsfunktion $\hat{F}$ eingeht. Erinnerung (vgl. Definition 3.1): Für $\mathbf{x} = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ ist für alle $z \in \mathbb{R}$

$$\hat{F}(z) = \hat{F}^{(x)}(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,z]}(x_i).$$

Die Konstruktion des Konfidenzbandes beruht auf folgender Ungleichung:

Satz 6.7 (Dvoretzky-Kiefer-Wolfowitz-Ungleichung)
Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariable mit Verteilungsfunktion F. Dann gilt für die assoziierte empirische Verteilungsfunktion $\hat{F}$ für alle $n = 1, 2, \ldots$ und alle $\varepsilon > 0$

$$\mathbb{P}\left(\sup_{z\in\mathbb{R}} \left|\hat{F}^{(\mathfrak{X})}(z) - F(z)\right| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

Für den Beweis siehe zum Beispiel van der Vaart (1998). Wir betonen, dass hier keinerlei Annahmen an die Verteilungsfunktion F gemacht werden und dass die Abschätzung für alle $n = 1, 2, \ldots$ gilt. Dies erlaubt sofort, ein Konfidenzband wie folgt zu konstruieren:

Korollar 6.8 (Konfidenzband)
Es seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariable mit $X_1 \sim \nu_\vartheta$, und ν_ϑ ist Mitglied der Familie $(\nu_\vartheta)_{\vartheta \in \Theta}$ aller reellen Verteilungen. Für $\alpha \in (0, 1)$ setze $\varepsilon_\alpha := (-\log(\alpha/2)/(2n))^{1/2}$. Dann bildet die folgende Statistik B, gegeben durch eine Menge von Intervallen

$$B(\mathbf{x}) = \left\{ \left[\hat{F}^{(\mathbf{x})}(z) - \varepsilon_\alpha, \hat{F}^{(\mathbf{x})}(z) + \varepsilon_\alpha \right] \middle| z \in \mathbb{R} \right\},$$

ein sogenanntes $(1 - \alpha)$-Konfidenzband für die empirische Verteilungsfunktion F_ϑ von ν_ϑ, denn es gilt für alle $n = 1, 2, \ldots$

$$\mathbb{P}_\vartheta \left(\bigcap_{z \in \mathbb{R}} \left\{ \left[\hat{F}^{(\mathfrak{X})}(z) - \varepsilon_\alpha, \hat{F}^{(\mathfrak{X})}(z) + \varepsilon_\alpha \right] \ni F_\vartheta(z) \right\} \right) > 1 - \alpha.$$

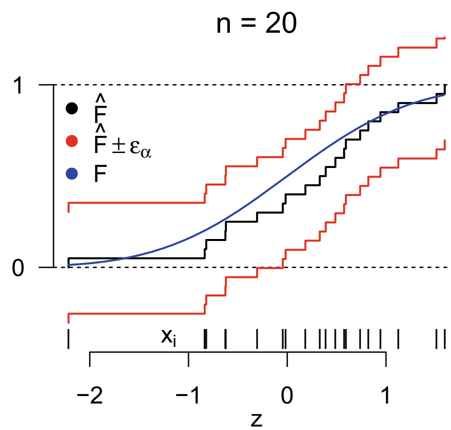
Bedeutung: Wenn ν_ϑ die wahre zugrunde liegende Verteilung ist, dann überdeckt das Intervall $[\hat{F}^{(\mathfrak{X})}(z) - \varepsilon_\alpha, \hat{F}^{(\mathfrak{X})}(z) + \varepsilon_\alpha]$ die assoziierte Verteilungsfunktion $F_\vartheta(z)$ – gleichmäßig an allen Stellen $z \in \mathbb{R}$ – mit großer Wahrscheinlichkeit.

Beweis Aus Satz 6.7 folgt aufgrund der Wahl von $\alpha = 2e^{-2n\varepsilon_\alpha^2}$

$$\mathbb{P}_\vartheta \left(\sup_{z \in \mathbb{R}} |\hat{F}(z) - F_\vartheta(z)| > \varepsilon_\alpha \right) \leq \alpha.$$

Abb. 6.5 zeigt ein Beispiel eines 95 %-Konfidenzbandes, konstruiert aus unabhängigen Realisierungen $x_1, \ldots, x_{20}$ aus der $N(0, 1)$-Verteilung. Das Konfidenzband (rot umrandeter Bereich) um die empirische Verteilungsfunktion $\hat{F}$ (schwarz) überdeckt die blaue Verteilungsfunktion F von $N(0, 1)$ vollständig. Der Zufall hat uns beim Ziehen der Beobachtungen also ein ‚moderates' Ereignis beschert; eines der Ereignisse, die im Sinne des Konfidenzbandes mit einer Wahrscheinlichkeit von mindestens 95 % auftreten. Wüssten wir nicht, dass die Beobachtungen aus der $N(0, 1)$-Verteilung stammen, so lieferten sie uns auch keinen Anlass, daran zu zweifeln. Eine mögliche Anwendung dieses Bandes wäre entsprechend, eine empirische Verteilung auf ihre Verträglichkeit mit einer angenommenen Verteilung zu untersuchen.

Abb. 6.5 Visualisierung des Konfidenzbandes



6.7 Der Satz von Glivenko und Cantelli

Im Kontext der Eigenschaften der empirischen Verteilungsfunktion wollen wir neben der Dvoretzky-Kiefer-Wolfowitz-Ungleichung aus Satz 6.7 noch ein anderes prominentes Resultat diskutieren, den Satz von Glivenko und Cantelli. Lassen wir im Kontext der Ungleichung $n \to \infty$ gehen, so folgern wir direkt, dass sich die empirische Verteilungsfunktion der wahren Verteilungsfunktion gleichmäßig annähert – und zwar mit großer Wahrscheinlichkeit. Der Satz von Glivenko und Cantelli besagt nun, dass diese gleichmäßige Annäherung sogar mit Wahrscheinlichkeit 1 gilt.

In der Praxis steht man häufig vor dem Problem, die Verteilung einer Statistik oder eines Schätzers ermitteln zu müssen. In Situationen, in denen man dazu analytisch nicht ohne Weiteres in der Lage ist, bieten Simulationen einen Ausweg: Durch mehrfache unabhängige Realisierung des Schätzers versucht man, seine wahre Verteilung zu approximieren. Dass das gut geht, besagt der Satz von Glivenko und Cantelli, welcher auch als Fundamentalsatz der Statistik bezeichnet wird. Als Hilfsresultat formulieren wir zunächst das folgende Lemma.

Lemma 6.9 (Punktweise Konvergenz der empirischen Verteilungsfunktion) *Es seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariable mit Verteilungsfunktion F. Für $n = 1, 2, \ldots$ sei $\mathfrak{X}_n = (X_1, \ldots, X_n)^t$. Dann gilt für alle $z \in \mathbb{R}$, dass für $n \to \infty$*

$$\hat{F}^{(\mathfrak{X}_n)}(z) \to F(z) \qquad \textit{mit Wahrscheinlichkeit } 1.$$

Beweis Die Aussage folgt direkt aus dem Starken Gesetz der großen Zahlen. Für alle $z \in \mathbb{R}$ gilt für $n \to \infty$

$$\hat{F}^{(\mathfrak{X}_n)}(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq z\}} \longrightarrow \mathbb{E}\left[\mathbb{1}_{\{X_1 \leq z\}}\right] = \mathbb{P}(X_1 \leq z) = F(z) \tag{6.10}$$

mit Wahrscheinlichkeit 1.

Dieses Resultat ist völlig allgemein gehalten. Wir stellen keine Annahmen an die Verteilungsfunktion F. Insbesondere sind die Voraussetzungen des Starken Gesetzes der Großen Zahlen erfüllt, denn die Indikatorvariablen sind nach Voraussetzung unabhängig und identisch $ber(F(z))$-verteilt.

Satz 6.10 (Satz von Glivenko und Cantelli)
Es seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariable mit Verteilungsfunktion F. Für $n = 1, 2, \ldots$ sei $\mathfrak{X}_n = (X_1, \ldots, X_n)^t$. Dann gilt für die assoziierte empirische Verteilungsfunktion $\hat{F}$ für $n \to \infty$

$$\sup_{z\in\mathbb{R}} \left|\hat{F}^{(\mathfrak{X}_n)}(z) - F(z)\right| \longrightarrow 0 \qquad \textit{mit Wahrscheinlichkeit 1.}$$

Bedeutung: Wenn wir nur genügend X_i beobachten, approximieren wir die wahre Verteilung auch beliebig genau – und zwar gleichmäßig. Wieder stellen wir keinerlei Annahmen an die Verteilungsfunktion F.

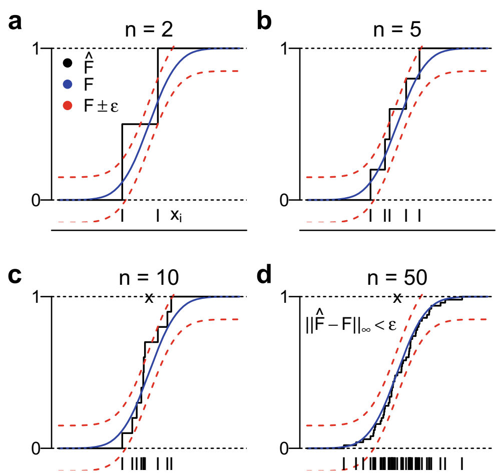


Abb. 6.6 Visualisierung des Satzes von Glivenko-Cantelli für $n \in \{2, 5, 10, 50\}$ (in **a, b, c, d**)

Abb. 6.6 visualisiert dieses Resultat am Beispiel der Verteilungsfunktion der $N(0, 1)$-Verteilung (blau) und eines ‚ε-Schlauchs' $[F - \varepsilon, F + \varepsilon]$ (rot, $\varepsilon = 0.15$). Die empirische Verteilungsfunktion $\hat{F}$ nähert sich mit wachsendem n der wahren Verteilungsfunktion F an. Bei $n = 50$ liegt $\hat{F}$ vollständig im ε-Schlauch.

Beweis von Satz 6.10 Wir wissen schon, dass für beliebiges, festes z gilt, dass $|\hat{F}(z) - F(z)| \to 0$ mit Wahrscheinlichkeit 1, vgl. Lemma 6.9, und damit konvergiert auch das Maximum über endlich viele z_i. Die Strategie ist daher, dass wir uns auf endlich viele Gitterpunkte $z_1, \ldots, z_k$ zurückziehen. Diese Gitterpunkte wählen wir so, dass auch die Abweichung $|F(z_i) - F(z_{i-1})|$ zwischen den Gitterpunkten klein bleibt.

Wir betrachten zunächst den Fall, dass F stetig ist. Für $\varepsilon > 0$ wählen wir $k = k(\varepsilon)$ Gitterpunkte $z_1, \ldots, z_k$ mit $-\infty = z_1 < z_2 < \ldots < z_k = \infty$, sodass

$$F(z_i) - F(z_{i-1}) \leq \frac{\varepsilon}{2} \tag{6.11}$$

für alle $i = 2, 3, \ldots, k$. Diese Zerlegung existiert, da F stetig und beschränkt ist. Jedes $z \in \mathbb{R}$ liegt nun in genau einem Intervall $[z_{i-1}, z_i)$, und es gilt

$$\begin{aligned}\hat{F}(z) - F(z) &\overset{(*)}{\leq} \hat{F}(z_i) - F(z_{i-1}) = [\hat{F}(z_i) - F(z_i)] + [F(z_i) - F(z_{i-1})] \\ &\overset{(6.11)}{\leq} \max_{i=1,\ldots,k} |\hat{F}(z_i) - F(z_i)| + \frac{\varepsilon}{2}. \end{aligned} \tag{6.12}$$

Anschaulich haben wir uns anhand der Monotonie von $\hat{F}$ und F in $(*)$ auf zwei der endlich vielen Gitterpunkte zurückgezogen und dann ausgenutzt, dass der Zuwachs von F zwischen den Gitterpunkten per Konstruktion durch $\varepsilon/2$ beschränkt ist. Mit analogen Argumenten folgt

$$\begin{aligned}\hat{F}(z) - F(z) &\overset{(*)}{\geq} \hat{F}(z_{i-1}) - F(z_i) = [\hat{F}(z_{i-1}) - F(z_{i-1})] + [F(z_{i-1}) - F(z_i)] \\ &\overset{(6.11)}{\geq} - \max_{i=1,\ldots,k} |\hat{F}(z_i) - F(z_i)| - \frac{\varepsilon}{2}. \end{aligned} \tag{6.13}$$

Da in (6.12) und (6.13) $z \in \mathbb{R}$ beliebig gewählt wurde und die ‚rechten Seiten' nicht mehr von z abhängen, können wir das betragsmäßige Supremum über *alle* z abschätzen via

$$\sup_{z\in\mathbb{R}} |\hat{F}(z) - F(z)| \leq \max_{i=1,\ldots,k} |\hat{F}(z_i) - F(z_i)| + \frac{\varepsilon}{2}.$$

Für $n \to \infty$ verschwindet nun das Maximum mit Wahrscheinlichkeit 1. Insbesondere finden wir das Maximum kleiner gleich $\varepsilon/2$ für n hinreichend groß mit Wahrscheinlichkeit 1, sodass das Supremum mit Wahrscheinlichkeit 1 kleiner ε wird. Da wir ε beliebig klein wählen können, folgt die Behauptung für stetiges F.

Der Fall von nichtstetigem F läuft über die gleichen Argumente: Ab einschließlich Ungleichung (6.11) ersetzen wir sämtliche Werte $F(z_i)$ und $\hat{F}(z_i)$ durch ihre ‚linken Limiten' $F(z_i^-) := \lim_{z \nearrow z_i} F(z)$ bzw. $\hat{F}(z_i^-) := \lim_{z \nearrow z_i} \hat{F}(z)$. Das hat zur Folge, dass wir trotz der Unstetigkeitsstellen wieder eine Zerlegung annehmen können, sodass (6.11) gilt. Insbesondere werden dabei Gitterpunkte in den Unstetigkeitsstellen von F liegen, wenn ε klein ist. Dann gehen wir analog vor, betrachten wieder ein Intervall $[z_{i-1}, z_i)$ und erhalten die Abschätzungen (6.12) und (6.13) wie oben und damit auch die Behauptung.

Bedeutung für die Statistik Wie angesprochen liegt die Bedeutung dieses Abschnittes darin, dass sich die Verteilung einer Statistik durch Simulation approximieren lässt. Wir machen den Wert dieser Aussage anhand eines Beispiels klar. In Abschn. 6.3 haben wir ein asymptotisches Konfidenzintervall I_n konstruiert, welches den unbekannten Erwartungswert mit gegebener Wahrscheinlichkeit zum Beispiel 5 % überdeckt. Aufgrund der Asymptotik ist dieses aber nur für große Stichprobengrößen n brauchbar. Was bedeutet groß?

Wir nehmen die praktisch untypische Sichtweise ein, die zugrunde liegende Verteilung zu kennen, und approximieren die Überdeckungswahrscheinlichkeit des Konfidenzintervalls für festes n durch Simulationen: Angenommen, unsere Beobachtungen seien durch unabhängige und $\gamma(\alpha, \lambda)$-verteilte Zufallsvariablen modelliert, mit $\alpha, \lambda > 0$, und zusammengefasst im Vektor $\mathfrak{X} = (X_1, \dots, X_n)^t$. Der Erwartungswert von X_1 ist $\mathbb{E}_{(\alpha,\lambda)}[X_1] = \alpha/\lambda$, und der Indikator $\mathbb{1}_{\{I(\mathfrak{X}) \ni \alpha/\lambda\}}$ gibt an, ob das Konfidenzintervall den Erwartungswert überdeckt. Wir argumentieren wie in Lemma 6.9 mit dem Starken Gesetz der großen Zahlen dafür, dass die relative Häufigkeit der Überdeckungen gegen die wahre Überdeckungswahrscheinlichkeit strebt

$$\frac{1}{k} \sum_{i=1}^{k} \mathbb{1}_{\{I(\mathfrak{X}^{(k)}) \ni \alpha/\lambda\}} \longrightarrow \mathbb{E}_{(\alpha,\lambda)}[\mathbb{1}_{\{I(\mathfrak{X}) \ni \alpha,\lambda\}}] = \mathbb{P}_{(\alpha,\lambda)}(I(\mathfrak{X}) \ni \alpha/\lambda)$$

mit Wahrscheinlichkeit 1, für $k \to \infty$, wenn $\mathfrak{X}^{(1)}, \mathfrak{X}^{(2)}, \dots$ unabhängige Kopien des Vektors $\mathfrak{X}$ beschreiben.

Für gegebene Parameter α und λ lässt sich eine Realisierung von $\mathfrak{X}^{(i)}$ nun schnell mit dem Rechner generieren, und für $k = 1000$ oder 10.000 wird die relative Häufigkeit der realisierten Überdeckungen eine vernünftige Approximation der Überdeckungswahrscheinlichkeit liefern. Dies kann man nun für verschiedene Stichprobengrößen n und abhängig von den Parametern α und λ untersuchen, wie in Abb. 6.3 dargestellt.

Die Frage der Überdeckungswahrscheinlichkeit spiegelt sich in $\{0, 1\}$-wertigen Indikatorvariablen wider, und wir kommen daher mit der punktweisen Argumentation aus Lemma 6.9 aus. Interessieren wir uns allgemeiner für die Verteilung einer reellwertigen Statistik, so ist die allgemeinere Aussage von Glivenko und Cantelli hilfreich. Beispielsweise besagt der Zentrale Grenzwertsatz, dass der Mittelwert unabhängiger Beobachtungen nach Reskalierung gegen die Normalverteilung strebt, wenn die Anzahl n der Beobachtungen zunimmt. Aber wie verteilt sich der Mittelwert bei fester Anzahl n? Vielleicht wissen wir, dass der

Mittelwert im Falle der Normalverteilung auch normalverteilt ist. Aber wie sieht es etwa bei der Gammaverteilung aus? Dazu können wir den Mittelwert einfach wieder zum Beispiel $k = 10.000$-mal realisieren und die empirische Verteilungsfunktion anschauen, welche dann gleichmäßig an der wahren Verteilungsfunktion des Mittelwertes liegen wird.