Satz 6.16. *$X_1, \dots, X_n$ u.i.v. $\sim \mathcal{N}_{\mu,\sigma^2}$ mit $\mu \in \mathbb{R}$, $\sigma > 0$,*

$$M := \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2.$$

$(X_i - M)_{i=1}^{n}$

Es gilt

1. M und S^2 sind unabhängig, $M \sim \mathcal{N}_{\mu,\sigma^2/n}$, $\frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$.

2. $T := \dfrac{\sqrt{n}(M-\mu)}{\sqrt{S^2}}$ ist Student-verteilt mit $n-1$ Freiheitsgraden.

$$\frac{\frac{\sqrt{n}}{\sigma}(M-\mu)}{\sqrt{S^2/\sigma^2}}$$

Bew.: O.b.d.A. sei $\mu = 0$, $\sigma^2 = 1$ (sonst betr. $\tilde{X}_i = \frac{X_i - \mu}{\sigma}$)

1. $(X_1, \dots, X_n)^T$ ist n-dim. Standardnormalverteilt.

Sei O orth. $n \times n$-Matrix mit erster Zeile $z_1 = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = O \cdot \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$ ist n-dim. Standardnormalvert. (Bsp 2.20)

Somit $Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i = \sqrt{n} \cdot M \quad (\sim \mathcal{N}_{0,1}) \Rightarrow M \sim \mathcal{N}_{0,1/n}$

$$(n-1)S^2 = \sum_{i=1}^{n} \underbrace{(X_i - M)^2}_{X_i^2 - 2MX_i + M^2} = \sum_{i=1}^{n} X_i^2 - n \cdot M^2 = \|(X_1, \dots, X_n)^T\|^2 - Y_1^2$$

$$= \cancel{Y_1^2} + \dots + Y_n^2 - \cancel{Y_1^2} \sim \chi^2_{n-1}$$

Korollar 6.17 (Student-Konfidenzintervall für den Erwartungswert im normalen Modell)**.** *Unter* P_ϑ, $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ *seien*

$$X_1, X_2, \dots, X_n \text{ u.i.v. } \sim \mathcal{N}_{\mu,\sigma^2}.$$

Sei $\alpha \in (0,1)$, $q = q_{n-1,1-\alpha/2}$ *das* $1 - \frac{\alpha}{2}$*-Quantil der Student-Verteilung mit* $n-1$ *Freiheitsgraden,*

$$M := \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2.$$

Dann ist

$$I := \left[M - q\sqrt{\frac{S^2}{n}}, M + q\sqrt{\frac{S^2}{n}}\,\right]$$

ein Konfidenzintervall für μ *zum Irrtumsniveau* α*.*

Beweis. $T := \frac{\sqrt{n}(M-\mu)}{\sqrt{S^2}}$ ist Student-verteilt mit $n-1$ Freiheitsgraden (für jede Wahl von μ und σ^2),

$$\begin{aligned}
&P_{(\mu,\sigma^2)}\Big(M - q\sqrt{\tfrac{S^2}{n}} \le \mu \le M + q\sqrt{\tfrac{S^2}{n}}\Big) \\
&= P(-q \le T \le q) = P(T \le q) - P(T \le -q) = 1 - \frac{\alpha}{2}.
\end{aligned}$$

□

Beispiel. Zwei Schlafmittel sollen verglichen werden, 10 Patienten erhielten in aufeinanderfolgenden Nächten Medikament A und B.

Die Daten (x_i = Anz. Stunden Schlaf mit Mittel A - Anz. Stunden Schlaf mit Mittel B bei Patient Nr. i):

i	1	2	3	4	5	6	7	8	9	10
x_i	1,2	2,4	1,3	1,3	0,0	1,0	1,8	0,8	4,6	1,4

Es ist

$$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i \approxeq 1{,}58, \qquad s = \Big(\frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2\Big)^{1/2} \approxeq 1{,}23.$$

Nehmen wir an, die Daten stammen aus einer Normalverteilung mit unbekanntem Mittelwert μ und unbekannter Varianz σ^2 (und die Ergebnisse der verschiedenen Patienten sind unabhängig).
Es ist $q_{9,\,0{,}995} \approxeq 3{,}25$ (aus einer Quantiltabelle oder beispielsweise mit R berechnet), demnach ist

$$\Big[\overline{x} \pm q\frac{s}{\sqrt{n}}\Big] \approxeq [0{,}31,\ 2{,}85]$$

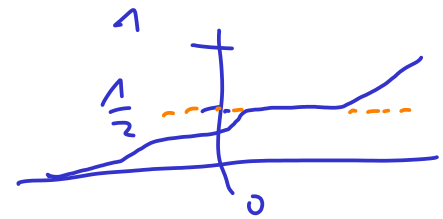
ein Konfidenzintervall für μ (die mittlere zusätzliche Anzahl Stunden Schlaf, die Medikament A mehr bringt als Medikament B) zum Sicherheitsniveau $0{,}99 = 1 - 0{,}01$.

Beachte: (Sinnlos) genaue Werte mit Rechnergenauigkeit sind $\overline{x} - q\frac{s}{\sqrt{n}} \approxeq 0{,}3159481$, $\overline{x} + q\frac{s}{\sqrt{n}} \approxeq 2{,}8440519$, man sollte allerdings die Grenzen eines Konfidenzintervalls stets „konservativ“, d.h. nach außen, runden.

6.3.2 Ein Konfidenzintervall für den Median (ein kleiner Ausflug in die nicht-parametrische Statistik)

Satz 6.18 (Ein Konfidenzintervall für den Median)**.** *Seien $X_1, \dots, X_n$ u.i.v. reellwertig, mit (unbekannter) Verteilung Q, die eine stetige Verteilungsfunktion besitzt (d.h. Q hat keine Atome).*
(Im Formalismus: $\Theta = \{\vartheta : \vartheta$ nicht-atomares W'maß auf $\mathbb{R}\}$, $\mathcal{X} = \mathbb{R}^n$, $P_\vartheta = \vartheta^{\otimes n}$)
$m(Q)$ sei „der" Median von Q (d.h. $Q\big((-\infty, m(Q)]\big) = \frac{1}{2} = Q\big([m(Q), \infty)\big)$, vgl. Def. 3.22; falls mehrere Werte in Frage kommen, nehmen wir das arithmetische Mittel aus dem kleinsten und dem größten möglichen Wert).
Die zugehörige Ordnungsstatistik ist

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}.$$



Zu $\alpha \in (0,1)$ wähle k maximal, so dass $\mathrm{Bin}_{n,1/2}(\{0, \dots, k-1\}) \le \frac{\alpha}{2}$, dann ist

$$\big[X_{(k)}, X_{(n-k+1)}\big]$$

ein Konfidenzintervall für den Median $m(Q)$ zum Sicherheitsniveau $1-\alpha$.

Beweis. Es ist

$$\begin{aligned} Q^{\otimes n}\big(X_{(k)} > m(Q)\big) &= Q^{\otimes n}\big(|\{1 \le i \le n : X_i \le m(Q)\}| \le k-1\big) \\ &= \mathrm{Bin}_{n,1/2}(\{0, \dots, k-1\}) \le \frac{\alpha}{2}, \end{aligned}$$

analog ist

$X_{(n-k+1)}$

$$Q^{\otimes n}\big(\cancel{X_{(n-k-1)}} < m(Q)\big) = Q^{\otimes n}\big(|\{1 \le i \le n : X_i \ge m(Q)\}| \le k-1\big) \le \frac{\alpha}{2},$$

somit

$X_{(n-k+1)}$

$$Q^{\otimes n}\Big(\big[X_{(k)}, X_{(n-k+1)}\big] \not\ni m(Q)\Big) \le Q^{\otimes n}\big(X_{(k)} > m(Q)\big) + Q^{\otimes n}\big(\cancel{X_{(n-k-1)}} < m(Q)\big) \le \alpha.$$

□

Nochmal die Daten aus dem „Schlafmittel-Vergleich“-Beispiel
(x_i = Anz. Stunden Schlaf mit Mittel A - Anz. Stunden Schlaf mit Mittel B bei Patient Nr. i):

i	1	2	3	4	5	6	7	8	9	10
x_i	1,2	2,4	1,3	1,3	0,0	1,0	1,8	0,8	4,6	1,4

Es ergäbe sich für $\alpha = 0{,}01$, $n = 10$:

Man muss in Satz 6.18 $k = 1$ wählen (es ist $\mathrm{Bin}_{10,1/2}(\{0\}) \approxeq 0{,}001$, aber $\mathrm{Bin}_{10,1/2}(\{0,1\}) \approxeq 0{,}012$), d.h. ein Konfidenzintervall für den Median (der Differenz der Schlafdauer unter Mittel A versus Mittel B) zum Sicherheitsniveau 99% ist $[X_{(1)}, X_{(10)}] = [0,\ 4{,}6]$.

(Für $\alpha = 0{,}05$ könnte man $k = 2$ wählen und erhielte $[X_{(2)}, X_{(9)}] = [0{,}8,\ 2{,}4]$ als Konfidenzintervall zum Sicherheitsniveau 95%.)

$$\mathrm{Bin}_{n,1/2}(\{0,1,\dots,k-1\}) \overset{!}{\leq} \frac{\alpha}{2}$$

6.3.3 Exkurs: Exakte Konfidenzintervalle für den Erfolgsparameter in der Binomialverteilung*

Unter n unabhängigen Versuchen seien x Erfolge beobachtet worden, wir fassen x als Realisierung einer $\mathrm{Bin}_{n,\vartheta}$-verteilten ZV auf und wollen anhand der Beobachtung auf ϑ schließen.

Wir hatten in Beispiel 6.12 das auf asymptotischer Normalität fußende (approximative) Konfidenzintervall für ϑ zum Niveau $1-\alpha$ kennen gelernt:

$$\left[\widehat{\vartheta} - q\frac{\widehat{\sigma}}{\sqrt{n}}, \widehat{\vartheta} + q\frac{\widehat{\sigma}}{\sqrt{n}}\right]$$

mit $\widehat{\vartheta} = \dfrac{x}{n}$, $\widehat{\sigma} = \sqrt{\widehat{\vartheta}(1-\widehat{\vartheta})}$, q das $1-\frac{\alpha}{2}$-Quantil von $\mathcal{N}_{0,1}$

Wie könnten wir vorgehen, wenn wir uns nicht auf die Asymptotik verlassen möchten? Wir beobachte $X \sim P_\vartheta := \mathrm{Bin}_{n,\vartheta}$ und möchten anhand der Beobachtung ein (nur nicht approximativ korrektes) Konfidenzintervall für $\vartheta \in \Theta = [0,1]$ konstruieren.

Idee: Zu $\vartheta \in \Theta := [0,1]$ wähle $c_\vartheta \in (0,1)$, so dass für

$$C_\vartheta := \left\{x \in \{0,1,\dots,n\} : \mathrm{Bin}_{n,\vartheta}(\{x\}) \geq c_\vartheta\right\}$$

gilt $\mathrm{Bin}_{n,\vartheta}(C_\vartheta) \geq 1-\alpha$ (und c_ϑ möglichst groß, so dass C_ϑ möglichst klein).

Setze $C(x) := \{\vartheta \in \Theta : x \in C_\vartheta\}$ für $x \in \mathcal{X} := \{0,1,\dots,n\}$, dann gilt

$$\forall\, \vartheta \in \Theta :\; P_\vartheta(\vartheta \in C(X)) = P_\vartheta(X \in C_\vartheta) \geq 1-\alpha$$

nach Konstruktion.

Es gilt

1. Für $\vartheta \in (0,1)$ ist $\{0,\dots,n\} \ni x \mapsto \mathrm{Bin}_{n,\vartheta}(\{x\})$ strikt wachsend auf $\{0,1,\dots,\lceil (n+1)\vartheta - 1\rceil\}$, strikt fallend auf $\{\lfloor (n+1)\vartheta \rfloor,\dots,n\}$, also maximal auf $x = \lfloor (n+1)\vartheta \rfloor$ (und auf $(n+1)\vartheta - 1$, wenn $(n+1)\vartheta \in \mathbb{Z}$).
2. Für $x \in \{1,\dots,n\}$ ist $[0,1] \ni \vartheta \mapsto Bin_{n,\vartheta}(\{x,x+1,\dots,n\})$ stetig, strikt monoton wachsend mit
$$\mathrm{Bin}_{n,\vartheta}(\{x,x+1,\dots,n\}) = \beta_{x,n-x+1}([0,\vartheta]),$$
wo $\beta_{a,b}$ (die Beta-Verteilung) die Dichte
$$f_{\mathrm{Beta}_{a,b}}(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1}$$
auf $(0,1)$ hat.

Argument:

1.
$$\begin{aligned}\frac{\mathrm{Bin}_{n,\vartheta}(\{x\})}{\mathrm{Bin}_{n,\vartheta}(\{x-1\})} &= \frac{\binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}}{\binom{n}{x-1}\vartheta^{x-1}(1-\vartheta)^{n-x+1}} \\ &= \frac{(n-x+1)\vartheta}{x(1-\vartheta)} > 1 \qquad \iff \quad x < (n+1)\vartheta\end{aligned}$$
2. $U_1,\dots,U_n$ unabhängig und uniform auf $[0,1]$, so ist
$$S_\vartheta := \sum_{i=1}^n \mathbf{1}_{[0,\vartheta]}(U_i)$$
ist $\mathrm{Bin}_{n,\vartheta}$-verteilt.

Sei $U_{(1)} < U_{(2)} < \ldots < U_{(n)}$ die „Ordnungsstatistik“.

$$\begin{aligned}
\mathrm{Bin}_{n,\vartheta}(\{x,\ldots,n\}) &= P(S_\vartheta \geq x) = P(U_{(x)} \leq \vartheta) \\
&= \sum_{k=1}^{n} \sum_{\substack{B\subseteq\{1,\ldots,n\}\setminus\{k\} \\ |B|=x-1}} \underbrace{P\begin{pmatrix} U_k \leq \vartheta, U_m \leq U_k \text{ für } m \in B, \\ U_l > U_k \text{ für } l \in \{1,\ldots,n\} \setminus (\{k\} \cup B) \end{pmatrix}}_{=\int_0^\vartheta u^{|B|}(1-u)^{n-|B|-1}\,du=\int_0^\vartheta u^{x-1}(1-u)^{n-x}\,du} \\
&= \underbrace{n\binom{n-1}{x-1}}_{=\frac{n!}{(x-1)!(n-x)!}=\frac{\Gamma(n+1)}{\Gamma(x)\Gamma(n-x+1)}} \int_0^\vartheta u^{x-1}(1-u)^{n-x}\,du
\end{aligned}$$

Wähle $C_\vartheta := \{x_-(\vartheta), x_-(\vartheta)+1, \ldots, x_+(\vartheta)\}$ mit $x_-(\vartheta) = \max\{x : \mathrm{Bin}_{n,\vartheta}(\{0,\ldots,x-1\}) \leq \frac{\alpha}{2}\}$ und $x_+(\vartheta) = \min\{x : \mathrm{Bin}_{n,\vartheta}(\{x+1,\ldots,n\}) \leq \frac{\alpha}{2}\}$.

Es gilt:

- $x \leq x_+(\vartheta) \iff \mathrm{Bin}_{n,\vartheta}(\{x,\ldots,n\}) = \beta_{x,n-x+1}([0,\vartheta]) > \frac{\alpha}{2}$
  $\iff \vartheta > p_-(x) := \frac{\alpha}{2}$-Quantil von $\mathrm{Beta}_{x,n-x+1}$.
- $x \geq x_+(\vartheta) \iff \mathrm{Bin}_{n,\vartheta}(\{0,\ldots,x\}) = 1 - \mathrm{Bin}(\{x+1,\ldots,n\}) = \beta_{x+1,n-x}([\vartheta,1]) \geq \frac{\alpha}{2}$.
  $\iff \vartheta < p_+(x) := 1-\frac{\alpha}{2}$-Quantil von $\mathrm{Beta}_{x+1,n-x}$.

Somit haben wir bewiesen:

Satz 6.19 (Exaktes Konfidenzintervall im Binomialmodell)**.**

$$p_-(x) := \tfrac{\alpha}{2}\text{-}\textit{Quantil von } \mathrm{Beta}_{x,n-x+1},$$
$$p_+(x) := 1 - \tfrac{\alpha}{2}\text{-}\textit{Quantil von } \mathrm{Beta}_{x+1,n-x}$$

$x \mapsto [p_-(x), p_+(x)]$ *ist ein Konfidenzintervall für* ϑ *zum Sicherheitsniveau* $1-\alpha$*.*

Bemerkung.

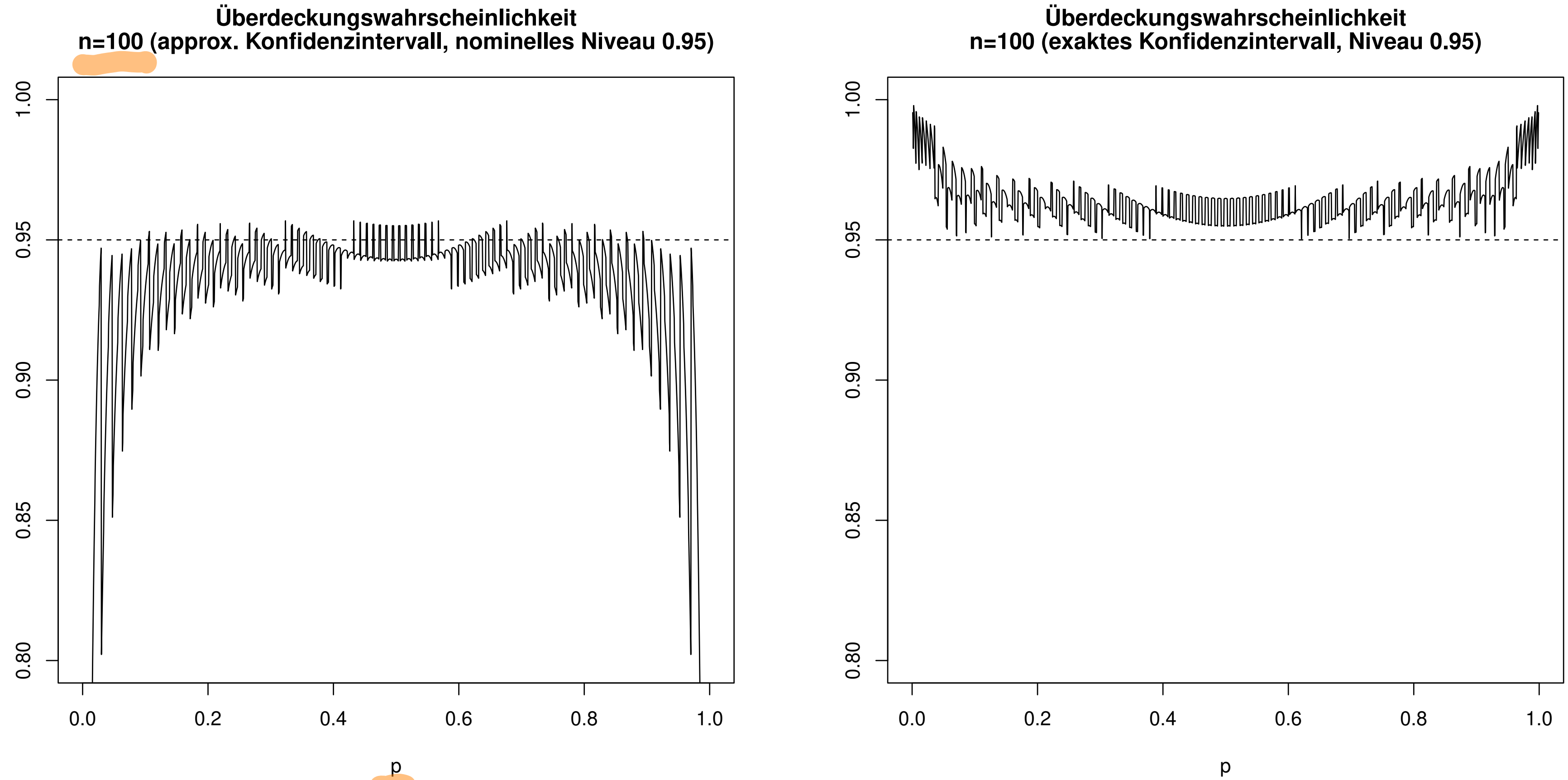
- Quantile der Beta-Verteilungen sind tabelliert, gelegentlich kann beim Nachschlagen in Tabellen die Symmetrieeigenschaft
$$\beta_{a,b}([0,x]) = \beta_{b,a}([1-x,1]) = 1 - \beta_{b,a}([0,1-x])$$
nützlich sein.

- `R` kennt die Beta-Verteilungen, ihre Verteilungsfunktionen `pbeta(x, a, b)` und ihre Quantile `qbeta(p, a, b)`

Für die Daten aus Beispiel 6.1 ($n = 53$, beobachtet $x = 23$) mit $\alpha = 0{,}05$ finden wir $\widehat{\vartheta} = \frac{23}{53} \approx 0{,}434$, $\widehat{\sigma} \approx 0{,}496$, $q_{0{,}975} \approx 1{,}96$, also ist das approximative 95%-Konfidenzintervall für ϑ hier $\left[\widehat{\vartheta} \pm q\frac{\widehat{\sigma}}{\sqrt{53}}\right] \approx [0{,}30, 0{,}57]$.

Es ist $p_-(23) = 0{,}025$-Quantil von $\beta_{23,31} \approx 0{,}30$, $p_+(23) = 0{,}975$-Quantil von $\beta_{24,30} \approx 0{,}57$, d.h. das exakte Konfidenzintervall $[p_-(x), p_+(x)] \approx [0{,}30, 0{,}57]$ (stimmt hier bei Rundung mit obigem überein).

(Die Intervalle sind nicht gleich: Absurd präzise Werte wären:
$\left[\widehat{\vartheta} \pm q\frac{\widehat{\sigma}}{\sqrt{53}}\right] \approx [0{,}3005306, 0{,}5673939]$, $[p_-(23), p_+(23)] \approx [0{,}2983921, 0{,}5771742]$.)



Überdeckungswahrscheinlichkeit eines (nominellen) 95%-Konfidenzintervalls für den Erfolgsparameter p einer Binomialverteilung mit $n = 100$ als Funktion von p: Links approximatives 95%-Konfidenzintervall (aus Bsp. 6.12), rechts exaktes Konfidenzintervall (aus Satz 6.19). Speziell für p nahe an 0 oder 1 hält das approximative Konfidenzintervall das nominelle Niveau nicht ein.

Bericht. Manchmal betrachtet man auch den Schätzer

$$\widetilde{\vartheta} = \frac{x+1}{n+2}$$

$\hat{\vartheta} = \frac{X}{n}$

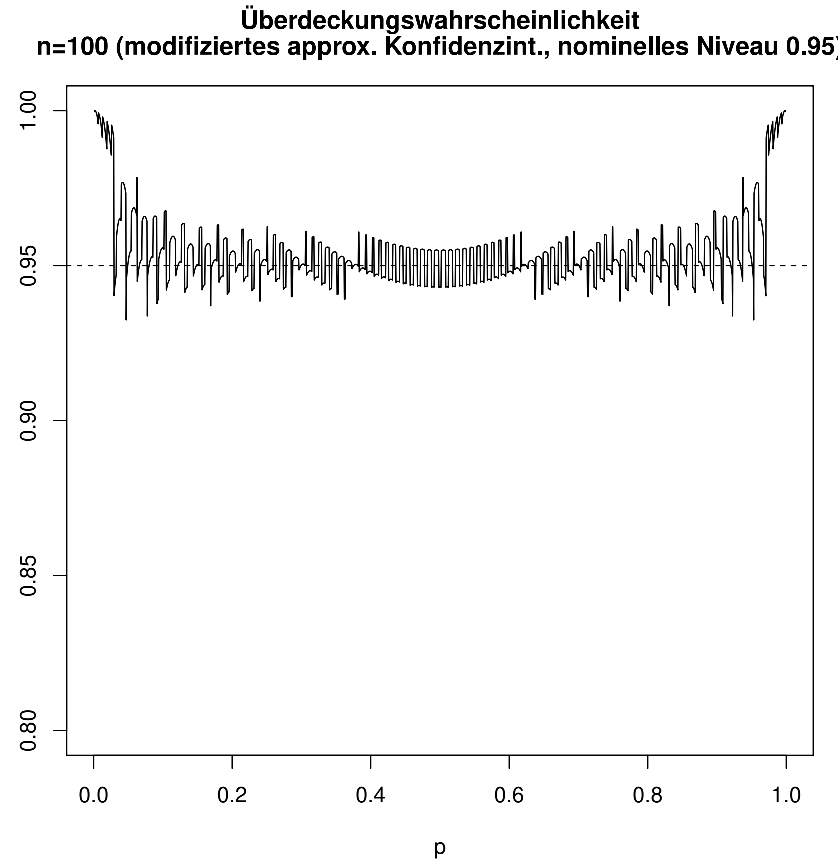
für ϑ und bildet als (approximatives) Konfidenzintervall zum Niveau $1-\alpha$

$$\Big[\widetilde{\vartheta} - q\frac{\widetilde{\sigma}}{\sqrt{n}}, \cancel{\widehat{\vartheta}} + q\frac{\widetilde{\sigma}}{\sqrt{n}}\Big]$$

$\widetilde{\vartheta}$

mit $\widetilde{\sigma} = \sqrt{\widetilde{\vartheta}(1-\widetilde{\vartheta})}$, q das $1-\frac{\alpha}{2}$-Quantil von $\mathcal{N}_{0,1}$.

$\widetilde{\vartheta}$ wirkt vielleicht zunächst „ad hoc" gewählt. Die Form von $\widetilde{\vartheta}$ ist natürlich im Kontext eines Bayesschen Ansatzes, den wir in dieser Vorlesung nicht weiter diskutieren werden.



6.4 Statistische Tests

Beispiel 6.20 (für einen einseitiger Binomialtest)**.** Herr A behauptet, (mit W'keit $\vartheta > 1/2$) vorhersagen zu können, ob die oberste Karte eines verdeckten, gut gemischten Skatblatts rot oder schwarz ist.

Frau B ist skeptisch (und verdächtigt, dass A einfach rät, d.h. $\vartheta = 1/2$) und schlägt vor, $n = 20$ Versuche durchzuführen.

Sei

$$X := \text{Anzahl richtige Vorhersagen von A},$$

wir modellieren $X \sim \mathrm{Bin}_{n,\vartheta}$ (mit uns unbekanntem $\vartheta \in [0,1]$).

$$X = \text{Anzahl richtige Vorhersagen von A unter } n = 20 \text{ Versuchen}$$

B wählt $\alpha = 0{,}05$, sagen wir, und k (möglichst klein) mit (und hier $\vartheta_0 := 1/2$)

$$\mathrm{Bin}_{n,\vartheta_0}\big(\{k, k+1, \dots, n\}\big) \le \alpha$$

(hier $k = 15$, denn $\mathrm{Bin}_{n,\vartheta_0}(\{15, 16, \dots, 20\}) \approxeq 0{,}021$, $\mathrm{Bin}_{n,\vartheta_0}(\{14, 15, \dots, 20\}) \approxeq 0{,}058$) und wird (auf dem Signifikanzniveau α) die

$$\text{Nullhypothese} \ : \ \vartheta \le \frac{1}{2}$$

verwerfen zugunsten der

$$\text{Alternative} \ : \ \vartheta > \frac{1}{2},$$

wenn das Ereignis $\{X \ge k\}$ eintritt, ansonsten die Nullhypothese beibehalten.

Demnach: Falls A tatsächlich rät (also in Wirklichkeit $\vartheta = 1/2$ gilt), ist die W'keit, ihm versehentlich hellseherische Fähigkeiten zuzuschreiben (dies wäre dann ein sogenannter „Fehler 1. Art“: die Nullhypothese abzulehnen, obwohl sie zutrifft), höchstens α.

Nehmen wir an, die 20 Versuche werden durchgeführt und A erzielt 13 „Treffer“. B wird also die Nullhypothese beibehalten (denn $\{X \ge k\}$ tritt nicht ein; quantitativer: der p-Wert ist $P_{1/2}(X \ge 13) \approxeq 0{,}132 > 0{,}05$; d.h., wenn A einfach nur rät, würde er in ca. 13,2% der Fälle mindestens ebensoviele „Treffer“ erzielen wie beobachtet) und etwa sagen:

> „Die Beobachtungen zeigen (auf dem Niveau $\alpha = 5\%$) keine keine signifikante Abweichung von der Nullhypothese.“

6.4.1 Der formale Rahmen statistischer Tests

Definition 6.21. Sei $(\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $\Theta = \Theta_0 \dot\cup \Theta_1$ disjunkte Zerlegung in „Nullhypothese“ und „Alternative“ (auch „Gegenhypothese“).

Eine Statistik $\varphi : \mathcal{X} \to [0,1]$ heißt ein *Test* von Θ_0 gegen Θ_1.

Der Test heißt *randomisiert*, wenn $\varphi(X) \notin \{0,1\}$, sonst *nicht-randomisiert*,

für einen nicht randomisierten Test φ heißt

$$\{x : \varphi(x) = 1\} \text{ der Ablehnungs- oder Verwerfungsbereich (von } \Theta_0).$$

$$G_\varphi : \Theta \to [0,1], \quad G_\varphi(\vartheta) = \mathbb{E}_\vartheta[\varphi]$$

heißt die *Gütefunktion* von φ ($1 - G_\varphi$ heißt Operationscharakteristisk), φ heißt ein Test zum Niveau (auch: Signifikanzniveau) $\alpha \in (0,1)$, wenn gilt

$$\sup_{\vartheta\in\Theta_0} G_\varphi(\vartheta) \le \alpha.$$

(Verwerfe Θ_0 mit W'keit $\varphi(x)$)

$$\sup_{\vartheta\in\Theta_0} G_\varphi(\vartheta) \quad \text{heißt effektives Niveau (auch: Umfang) von } \varphi.$$

Für $\vartheta \in \Theta_1$ heißt $G_\varphi(\vartheta)$ die Macht (auch: Schärfe, englisch: power) des Tests φ bei ϑ.

Sei $(\varphi_\alpha)_{\alpha\in(0,1)}$ eine Familie von nicht-randomisierten Tests mit $\varphi_\alpha \le \varphi_{\alpha'}$ für $\alpha \le \alpha'$, φ_α habe effektives Niveau α. Dann heißt für $x \in \mathcal{X}$

$$p\,\big(= p(x)\big) = \inf\{\alpha \in (0,1) : \varphi_\alpha(x) = 1\}$$

der *p-Wert* (bei Beobachtung x).

Interpretation.

1. Man interpretiert φ als Entscheidungsregel: Bei gegebener Beobachtung x
$$\begin{aligned} \varphi(x) = 0 &\ : \ \text{behalte Nullhypothese bei} \\ \varphi(x) = 1 &\ : \ \text{verwirf Nullhypothese, entscheide für die Alternative} \\ \varphi(x) \in (0,1) &\ : \ \begin{array}{l}\text{wirf eine Münze, die mit W'keit } \varphi(x) \\ \text{für die Alternative entscheidet}\end{array} \end{aligned}$$
2. Niveau α bedeutet, dass die W'keit für einen „Fehler 1. Art“ (die Nullhypothese fälschlicherweise zu verwerfen) $\leq \alpha$ ist (uniform in $\vartheta \in \Theta_0$).
3. Für $\vartheta \in \Theta_1$ ist $1 - G_\varphi(\vartheta)$ die Wahrscheinlichkeit, einen „Fehler 2. Art“ zu begehen (die Nullhypothese fälschlicherweise zu akzeptieren).
4. Viele „praktische“ Tests haben die folgende Form, z.B. Bspe. 6.23, 6.24, 6.25, 6.32: Berechne eine gewisse (Test-)Statistik Y (aus den Beobachtungen), verwirf die Nullhypothese, wenn $Y > q$ für einen gewissen Wert $q = q(\alpha)$, der in Abhängigkeit von den Parametern des Tests (insbesondere dem gewünschten Niveau α) gewählt wird. In der Sprache von Definition 6.21 also: $\varphi_\alpha(x) = \mathbf{1}(Y(x) > q(\alpha))$ und $\sup_{\vartheta\in\Theta_0} \mathbb{E}_\vartheta[\varphi] = \alpha$.

   Dann kann man den p-Wert des Test(ergebnisses) interpretieren als die Wahrscheinlichkeit, bei Gültigkeit der Nullhypothese einen mindestens so „extremen“ Wert der Teststatistik zu finden wie den tatsächlich anhand der Daten beobachteten.

Demnach sind für φ wünschenswert:

$$G_\varphi \text{ sollte auf } \Theta_1 \text{ möglichst groß sein}$$

(solange mit dem gewünschten Signifikanzniveau verträglich), zudem sollte für einen Test zum Niveau α gelten

$$\sup_{\vartheta\in\Theta_0} G_\varphi(\vartheta) \leq \alpha \leq \sup_{\vartheta\in\Theta_1} G_\varphi(\vartheta) \qquad \text{(dann heißt } \varphi \text{ „unverfälscht“).}$$