

## Tests: der p-Wert

Sei  $X_1, \dots, X_n$  eine Stichprobe vom Umfang  $n$ . Wir wollen eine Hypothese  $H_0 : \vartheta = \vartheta_0$  gegen eine Alternative  $H_A : \vartheta \in \Theta_A$  testen. Wir nennen  $T = t(X_1, \dots, X_n)$  die Teststatistik und  $K$  den kritischen Bereich (zum fixierten Niveau  $\alpha$ ). Seien  $x_i = X_i(\omega)$ ,  $i = 1, \dots, n$ , die realisierten Daten. Bei einem Test zum Niveau  $\alpha$  wird dann also die Hypothese abgelehnt, wenn  $T(\omega) = t(x_1, \dots, x_n)$  einen Wert in  $K$  ergibt, und wir haben  $P_{\vartheta_0}[T \in K] \leq \alpha$ .

Alternativ zu diesem Vorgehen kann man auch den sogenannten *p-Wert* berechnen; das ist die Wahrscheinlichkeit, unter der Hypothese  $H_0$  einen mindestens so extremen Wert für  $T$  zu bekommen wie den Wert aus der Stichprobe. “Extrem” bezieht sich dabei auf die Alternative — ist also beispielsweise  $H_A : \vartheta > \vartheta_A$  und entspricht das einem grossen Wert von  $T$ , so ist

$$\text{p-Wert}(\omega) = P_{\vartheta_0}[T > t_0] \Big|_{t_0=t(X_1(\omega), \dots, X_n(\omega))},$$

d.h. wir berechnen die Wahrscheinlichkeit  $t_0 \mapsto P_{\vartheta_0}[T > t_0]$  als Funktion von  $t_0$  und setzen dann für  $t_0$  den zufälligen Wert  $T(\omega)$  ein. Analog berechnet man den p-Wert bei einer Alternative der Form  $H_A : \vartheta < \vartheta_A$  oder der Form  $H_A : \vartheta \neq \vartheta_0$ .

Wie die Stichprobe ist auch der p-Wert eine Zufallsvariable (!), weil er von den Daten abhängt. Konzeptuell haben wir das Niveau  $\alpha$  und damit auch den kritischen Bereich  $K$  fixiert, und diese hängen im Gegensatz zum p-Wert nicht von den Daten ab. Der p-Wert hat immer Werte in  $[0, 1]$ , und man kann zeigen, dass er unter  $P_{\vartheta_0}$  auf  $[0, 1]$  gleichverteilt ist.

Die Entscheidung beim Test ist dann, dass man die Hypothese ablehnt, sofern der beobachtete p-Wert kleiner als das Niveau  $\alpha$  ist.

Ein Vorteil des p-Wertes ist, dass viele Statistik-Pakete direkt einen p-Wert berechnen. Zudem hat man etwas mehr Information als nur die reine Testentscheidung — der p-Wert gibt eine Indikation dafür, wie weit der Wert der Teststatistik im kritischen Bereich liegt oder nicht.

Umgekehrt muss man mit der Interpretation des p-Wertes mehr als bei einem Test auf-

passen; beispielsweise ist die “Aussage”

“der p-Wert ist die Wahrscheinlichkeit, dass die Hypothese richtig ist”

völlig falsch, denn der p-Wert ist eine Zufallsvariable, während die Hypothese mit Sicherheit entweder richtig oder falsch ist (wir wissen einfach nicht, welches von beidem der Fall ist). Zudem kann man mit unüberlegtem oder systematischem Wiederholen von Experimenten den p-Wert deutlich verfälschen.