Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)

Lehrstuhl für Ökonometrie und Statistik

Kapitel VIII - Tests zum Niveau α

Induktive Statistik

Prof. Dr. W.-D. Heller
Hartwig Senska
Carlo Siebenschuh

Testsituationen sind dadurch charakterisiert, dass genau zwei Entscheidungen möglich sind.

Beispiele:

- Annahme oder Ablehnung einer Warenpartie
- Eingriff in die Produktion oder nicht
- Behandlung mit einem bestimmten Medikament oder nicht
- Durchführung einer Investitionsentscheidung oder nicht
- Anmeldung zu einer Prüfung oder nicht

Die Entscheidung wird bei einer parametrischen Verteilungsannahme explizit oder implizit über den Parameter der Verteilung der Zufallsvariable Y (Umweltzustand der 1. Grundannahme) getroffen.

Konsequenz:

Der Parameterraum Γ ist aufgeteilt in zwei Bereiche Γ_0 und Γ_1:

$$\Gamma = \Gamma_0 \cup \Gamma_1 \quad \text{mit} \quad \Gamma_0 \neq \emptyset \neq \Gamma_1 \quad \text{und} \quad \Gamma_0 \cap \Gamma_1 = \emptyset.$$

Beispiele:

1. Warenpartie: $\Gamma_0 = [0, p_0]$, $\Gamma_1 = (p_0, 1]$
2. Justierung einer Anlage: $\Gamma_0 = \{\mu_0\}$, $\Gamma_1 = \Gamma \backslash \{\mu_0\}$
   (μ_0 Sollwert: $\Gamma_0 : \mu = \mu_0$ $\Gamma_1 : \mu \neq \mu_0$)

Zwei Hypothesen über den wahren Parameter γ:

$$H_0 : \gamma \in \Gamma_0 \qquad\qquad H_1 : \gamma \in \Gamma_1$$

"Nullhypothese" *"Gegenhypothese"*

Information als Entscheidungsgrundlage:

Ergebnis $(x_1, \ldots, x_n)$ einer Stichprobe vom Umfang n mit Zurücklegen.

Die Entscheidung hängt ab vom Stichprobenergebnis:

$\rightarrow$ Entscheidungsfunktion $\delta : \mathcal{X} \rightarrow \{d_0, d_1\}$

- $d_1 : H_1$ wird als richtig angesehen
- $d_0 : H_0$ wird als richtig angesehen

  oder schwächer

  H_0 wird nicht abgelehnt

$\delta(x_1, \ldots, x_n)$ Entscheidung beim Stichprobenergebnis $(x_1, \ldots, x_n)$

$X = (X_1, \ldots, X_n)$ Stichprobenvektor

$\delta(X_1, \ldots, X_n)$ Funktion eines Zufallsvektors mit den Werten d_0 und d_1

$\delta(X_1, \ldots, X_n)$ wird Zufallsvariable, falls d_0 und d_1 durch Zahlen ersetzt (kodiert) werden. Üblich ist 0 für d_0 und 1 für d_1, $\delta(X_1, \ldots, X_n)$ ist dann Bernoulli-verteilt.

Die Entscheidungsfunktion δ hat nur zwei Werte

$\hookrightarrow$ Aufteilung des Stichprobenraums in zwei Teilbereiche

Annahmebereich: Wert d_0

$$A_\delta = \{x \in \mathcal{X} | \delta(x) = d_0\}$$

Ablehnungsbereich, Kritischer Bereich: Wert d_1

$$K_\delta = \{x \in \mathcal{X} | \delta(x) = d_1\}$$

Jede Entscheidungsfunktion entspricht einer Aufteilung und umgekehrt.

Die Information ist unvollständig. Daher sind Fehlentscheidungen möglich:

		γ aus	
		Γ_0	Γ_1
Entscheidung	d_0	korrekt	Fehler 2. Art
	d_1	Fehler 1. Art	korrekt

- Fehler 1. Art: H_0 wird fälschlicherweise abgelehnt
- Fehler 2. Art: H_0 wird fälschlicherweise angenommen

(Die Schadensfunktion wird sich bei den vier angegebenen Konstellationen von Entscheidung und Parameterteilbereich wesentlich unterscheiden.)

Konstanter Schadensverlauf in den Teilbereichen Γ_0 und Γ_1 des Parameterraums führt zu:

Konstante Schadensfunktion:

$$S(d,\gamma) = \begin{cases} c_1 & \gamma \in \Gamma_0, d = d_1 & \text{“Fehler 1. Art”} \\ c_2 & \gamma \in \Gamma_1, d = d_0 & \text{“Fehler 2. Art”} \\ 0 & \text{sonst} & \text{korrekte Entscheidung} \end{cases}$$

Die konstante Schadensfunktion ist grob und undifferenziert, da der Schaden beim Fehler 1. Art nicht von $\gamma \in \Gamma_0$ und beim Fehler 2. Art nicht von $\gamma \in \Gamma_1$ abhängt.

Beispiel: Warenpartie

Fehler 1. Art: Ablehnung der Warenpartie, obwohl $p \leq p_0$.
Der Schaden hängt sicher auch davon ab, ob p nahe bei p_0 *oder nahe bei 0 ist*.

Fehler 2. Art: Annahme der Warenpartie, obwohl $p > p_0$.
Der Schaden ist offensichtlich auch davon abhängig, wie groß der Ausschussanteil wirklich ist.

Lineare Schadensfunktion:

$$S(p,d) = \begin{cases} c_1(p_0 - p) & p \in [0, p_0], d = d_1 \\ c_2(p - p_0) & p \in (p_0, 1], d = d_0 \\ 0 & \text{sonst} \end{cases}$$

oder quadratische Schadensfunktion:

$$S(p,d) = \begin{cases} c_1(p_0 - p)^2 & p \in [0, p_0], d = d_1 \\ c_2(p - p_0)^2 & p \in (p_0, 1], d = d_0 \\ 0 & \text{sonst} \end{cases}$$

Risikofunktion bei konstanter Schadensfunktion:

$$
\begin{aligned}
R(\delta,\gamma) &= E_\gamma[S(\delta(X),\gamma)] \\
&= S(d_0,\gamma)P_\gamma(\delta(X)=d_0)+S(d_1,\gamma)P_\gamma(\delta(X)=d_1)
\end{aligned}
$$

$$
= \begin{cases} 0 \quad P_\gamma(\delta(X)=d_0)+c_1P_\gamma(\delta(X)=d_1) & \text{für} \quad \gamma\in\Gamma_0 \\ \\ c_2 \quad P_\gamma(\delta(X)=d_0)+0P_\gamma(\delta(X)=d_1) & \text{für} \quad \gamma\in\Gamma_1. \end{cases}
$$

Wesentliche Bestandteile der Risikofunktion sind also die Fehlerwahrscheinlichkeiten:

$P_\gamma(\delta(X) = d_1)$ für $\gamma \in \Gamma_0$
Wahrscheinlichkeit für den Fehler 1. Art
(kurz: **Fehlerwahrscheinlichkeit 1. Art bei** δ : $P_I(\delta, \gamma)$)

$P_\gamma(\delta(X) = d_0)$ für $\gamma \in \Gamma_1$
Wahrscheinlichkeit für den Fehler 2. Art
(kurz: **Fehlerwahrscheinlichkeit 2. Art bei** δ : $P_{II}(\delta, \gamma)$)

Achtung:
Die Fehlerwahrscheinlichkeit variiert mit $\gamma \in \Gamma_0$ bzw. $\gamma \in \Gamma_1$, ist also nicht konstant in Γ_0 bzw. Γ_1.

Ziel: Entscheidungsfunktion mit minimalen Fehlerwahrscheinlichkeiten für alle γ.

$$\begin{aligned} \delta(x) = d_1 &\Leftrightarrow x \in K_\delta \\ \Rightarrow P_\gamma(\delta(X) = d_1) &= P_\gamma(X \in K_\delta) \\ \delta(x) = d_0 &\Leftrightarrow x \in A_\delta \\ \Rightarrow P_\gamma(\delta(X) = d_0) &= P_\gamma(X \in A_\delta) \end{aligned}$$

Daher im Allgemeinen:

$$\begin{array}{ccc} \textit{Verkleinerung der} & & \textit{Vergrößerung der} \\ \textit{Fehlerwahrscheinlichkeit} & \Longleftrightarrow & \textit{Fehlerwahrscheinlichkeit} \\ \textit{1. Art} & & \textit{2. Art} \\ & & \\ \Updownarrow & & \Updownarrow \\ & & \\ \textit{Verkleinerung von } K_\delta & \Longleftrightarrow & \textit{Vergrößerung von } A_\delta \end{array}$$

Trade-Off zwischen den Fehlerwahrscheinlichkeiten 1. und 2. Art.

Beispiel für eine Ausnahme:

Servicezeit Y gleichverteilt auf $[0, \gamma]$ mit $\gamma = \gamma_0 = 10$h oder $\gamma = \gamma_1 = 15$h

Mit einer Stichprobe soll getestet werden, ob

$$H_0 : \gamma = \gamma_0 = 10 \text{ oder } H_1 : \gamma = \gamma_1 = 15$$

als richtig angesehen wird.

$T(x_1, \ldots, x_n) = max\{x_1, \ldots, x_n\}$ ist suffiziente Statistik bezüglich γ. Daher ist es sinnvoll, die Entscheidung vom Maximalwert $T(x_1, \ldots, x_n)$ der Stichprobe abhängig zu machen. Ein großer Maximalwert spricht für $\gamma = \gamma_1 = 15$, ein kleiner für $\gamma = \gamma_0 = 10$.

Entscheidungsfunktion:

$$\delta_c(x_1, \ldots, x_n) = \begin{cases} d_0 & max\{x_1, \ldots, x_n\} \leq c \\ d_1 & max\{x_1, \ldots, x_n\} > c \end{cases}$$

mit einer geeigneten **Testschranke** c.

Die Wahl von c beeinflusst die Fehlerwahrscheinlichkeiten 1. und 2. Art.

Fehlerwahrscheinlichkeit 1. Art: Gegeben $\gamma = \gamma_0 = 10$

$$P_{\gamma_0}(T(X) > c) = 1 - P_{\gamma_0}(T(X) \leq c) = 1 - F_{T(X),\gamma_0}(c)$$

Für die Verteilungsfunktion des Maximalwerts gilt:

$$F_{T(X),\gamma_0}(t) = (F_{Y,\gamma_0}(t))^n = \begin{cases} \frac{1}{\gamma_0^n} t^n & \text{für } 0 \leq t \leq \gamma_0 \\ 1 & \text{für } \gamma_0 \leq t \end{cases}$$

Damit:

$$P_I(\delta, \gamma_0) = P_{\gamma_0}(T(X) > c) = \begin{cases} 1 - \frac{1}{\gamma_0^n} c^n & 0 \leq c \leq \gamma_0 \\ 0 & \gamma_0 \leq c \end{cases}$$

Die Fehlerwahrscheinlichkeit 1. Art ist für $c \geq \gamma_0$ Null, da bei Gültigkeit der Nullhypothese ein Maximalwert ≥ 10 nicht vorkommen kann. Ab $c = 10$ wirkt sich eine Vergrößerung des Annahmebereichs beim Fehler 1. Art nicht mehr aus. Wohl aber bei der

Fehlerwahrscheinlichkeit 2. Art: Gegeben $\gamma = \gamma_1 = 15$

$$P_{II}(\delta, \gamma_1) = P_{\gamma_1}(T(X) \leq c) = F_{T(X),\gamma_1}(c) = \begin{cases} \frac{1}{\gamma_1^n} c^n & 0 \leq c \leq \gamma_1 \\ 1 & \gamma_1 \leq c \end{cases}$$

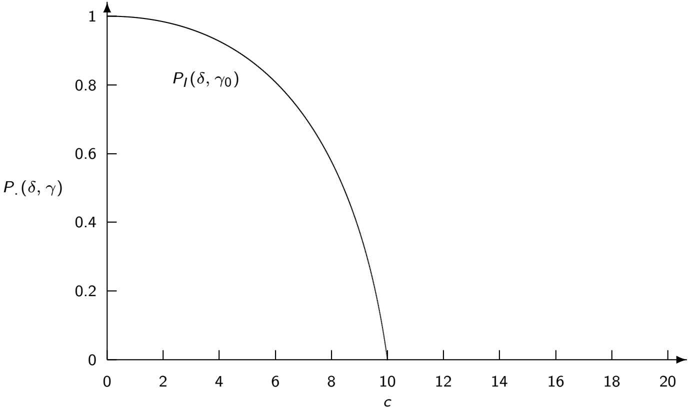

Abbildung: Verlauf der Fehlerwahrscheinlichkeiten in Abhängigkeit von c ($n = 5$).

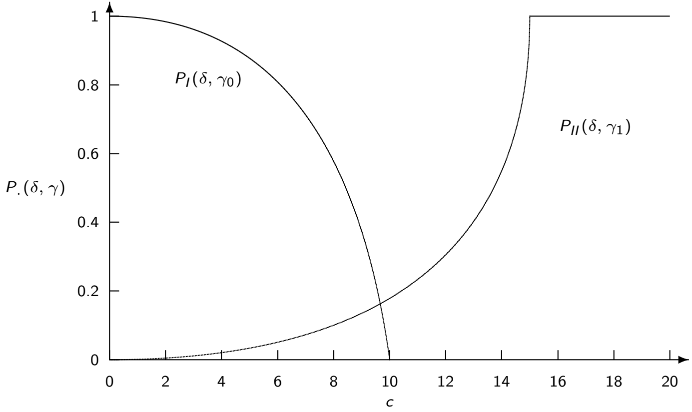

Abbildung: Verlauf der Fehlerwahrscheinlichkeiten in Abhängigkeit von c ($n = 5$).

$0 \leq c < 10$:	Eine Vergrößerung von c verkleinert die Fehlerwahrscheinlichkeit 1. Art und vergrößert die Fehlerwahrscheinlichkeit 2. Art (entspricht dem "Normalfall").
$10 \leq c \leq 15$:	Eine Vergrößerung von c vergrößert die Fehlerwahrscheinlichkeit 2. Art. Fehlerwahrscheinlichkeit 1. Art $= 0$, kann also nicht weiter reduziert werden.
Folgerung:	$c > 10$ ist als Testschranke nicht sinnvoll!

Das Beispiel zeigt auch, **dass der Stichprobenraum vom Parameter abhängen kann**:

- Bei $\gamma = \gamma_0 = 10$ ist der Stichprobenraum $\{(x_1, \ldots, x_n) | x_i \in [0, 10]\}$
- Bei $\gamma = \gamma_1 = 15$ ist der Stichprobenraum $\{(x_1, \ldots, x_n) | x_i \in [0, 15]\}$

Da der Parameter und damit der Stichprobenraum nicht bekannt ist, ist die Entscheidungsfunktion für den umfangreicheren Stichprobenraum festzulegen.

Folgerung aus dem Trade-Off zwischen den Fehlerwahrscheinlichkeiten: Gemeinsame Minimierung der beiden Fehlerwahrscheinlichkeiten ist nicht möglich.

Die Beurteilung einer Entscheidungsfunktion erfolgt durch die Analyse des Verlaufs der Wahrscheinlichkeiten

$P_\gamma(\delta(X) = d_0)$ oder $P_\gamma(\delta(X) = d_1) = 1 - P_\gamma(\delta(X) = d_0)$

in den Bereichen Γ_0 und Γ_1.

Definition:
Die Funktion L_δ mit

$$L_\delta(\gamma) = P_\gamma(\delta(X) = d_0) \text{ für } \gamma \in \Gamma$$

heißt **OC-Funktion** oder **Operations-Charakteristik** der Entscheidungsfunktion δ.

$L_\delta(\gamma)$ ist die Annahmewahrscheinlichkeit der Nullhypothese (Nichtablehnwahrscheinlichkeit der Nullhypothese) bei der Entscheidungsfunktion δ **in Abhängigkeit** von γ; für $\gamma \in \Gamma_0$ also die Wahrscheinlichkeit für eine korrekte Entscheidung, bei $\gamma \in \Gamma_1$, die Wahrscheinlichkeit für den Fehler 2. Art.

Definition:
Die Funktion G_δ mit

$$G_\delta(\gamma) = P_\gamma(\delta(X) = d_1) \text{ für } \gamma \in \Gamma$$

heißt **Gütefunktion (Power)** der Entscheidungsfunktion δ.

$G_\delta(\gamma)$ ist die Ablehnwahrscheinlichkeit der Nullhypothese (Annahme- wahrscheinlichkeit der Gegenhypothese) bei der Entscheidungsfuntion δ **in Abhängigkeit** von γ; bei $\gamma \in \Gamma_0$ also die Fehlerwahrscheinlichkeit 1. Art, bei $\gamma \in \Gamma_1$ die Wahrscheinlichkeit für eine korrekte Entscheidung.

Es gilt: $L_\delta(\gamma) = 1 - G_\delta(\gamma)$

Zielsetzung bzgl. der OC-Funktion:

- Werte nahe bei 1 für $\gamma \in \Gamma_0$
- Werte nahe bei 0 für $\gamma \in \Gamma_1$

Trade-Off zwischen den Fehlerwahrscheinlichkeiten 1. und 2. Art bedeutet: Bei einer "guten" Entscheidungsfunktion ist es nicht möglich, die OC-Funktion im Bereich Γ_0 zu erhöhen **und** im Bereich Γ_1 zu senken.

Neue Zielsetzung:

- Einhaltung einer Schranke für die Fehlerwahrscheinlichkeiten 1. Art
- Minimierung der Fehlerwahrscheinlichkeiten 2. Art bei Einhaltung der Schranke

Definition: Test zum Niveau α:
Ein Test $\delta : \mathcal{X} \to \{d_0, d_1\}$ heißt *Test zum Niveau* α, $0 < \alpha < 1$, falls

$$P_I(\delta, \gamma) = P_\gamma(\delta(X) = d_1) \leq \alpha \quad \text{für alle} \quad \gamma \in \Gamma_0$$

gilt.

Definition:
Ein Test δ zum Niveau α heißt *gleichmäßig bester Test zum Niveau* α, falls für jeden Test δ' zum Niveau α

$$P_{II}(\delta,\gamma) = P_\gamma(\delta(X) = d_0) \leq P_{II}(\delta',\gamma) = P_\gamma(\delta'(X) = d_0) \;\; \textbf{für alle} \;\; \gamma \in \Gamma_1$$

gilt.

Übliche Werte für α : $0.1, 0.05, 0.01$

Entscheidet man sich bei einem Test zum Niveau α für d_1, so gilt die Hypothese H_1 als "statistisch" gesichert: Denn die Entscheidung ist nur falsch, wenn H_0 richtig ist.
Bei einem Test zum Niveau α tritt dieser Fehler 1. Art aber nur mit einer Wahrscheinlichkeit von höchstens α auf. Bei den üblichen Werten von α also mit kleiner Wahrscheinlichkeit.

Aber:

Die Aussage "d_1 ist mit Wahrscheinlichkeit 1-α richtig" ist unsinnig, denn d_1 ist entweder richtig oder falsch, da der vorliegende wahre Parameterwert nicht zufällig ist. d_1 wird lediglich mit einer Wahrscheinlichkeit gewählt oder nicht.

Also:
Die Verwendung eines Tests zum Niveau α besagt nur, dass bei diesem Entscheidungsverfahren der Fehler 1. Art mit geringer Wahrscheinlichkeit auftritt. Über die Wahrscheinlichkeit bei Gültigkeit der Gegenhypothese die Entscheidung d_1 (bzw. d_0) zu treffen, sagt die Eigenschaft "zum Niveau α" nichts aus. Dazu muss der Verlauf der Gütefunktion (bzw. der OC-Funktion) betrachtet werden.

Daher sollte das Testproblem immer so formuliert werden, dass der Fehler 1. Art die schwerwiegenderen Folgen hat. Die Wahrscheinlichkeit für diesen Fehler wird bei einem Test zum Niveau α kontrolliert.

Die Wahrscheinlichkeit für den Fehler 2. Art kann bis zum Wert $1 - \alpha$ ansteigen. Daher wird d_0 in folgender Weise interpretiert:

d_0 : H_0 **kann nicht abglehnt werden**

Das **Problem bei Tests** ist die **Abhängigkeit der Fehlerwahrscheinlichkeit 1. Art und 2. Art vom Parameter** γ.

Daher werden zunächst "einfache" Testprobleme behandelt, bei denen

der Parameter γ nur einen von zwei Werten γ_0 und γ_1 annehmen kann, der Parameterraum besteht also aus zwei Elementen:

$$\Gamma = \{\gamma_0, \gamma_1\}$$

Damit lauten die Hypothesen

$$H_0 : \gamma = \gamma_0 \qquad\qquad H_1 : \gamma = \gamma_1$$

Die Kombinationsmöglichkeiten aus Entscheidung und Parameter reduzieren sich auf vier Kombinationen:

$$(d_0, \gamma_0), (d_0, \gamma_1), (d_1, \gamma_0), (d_1, \gamma_1)$$

mit

	$\gamma = \gamma_0$	$\gamma = \gamma_1$
d_0	korrekte E.	Fehler 2. Art
d_1	Fehler 1. Art	korrekte E.