Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)

Lehrstuhl für Ökonometrie und Statistik

Kapitel X - Gleichmäßig beste Tests zum Niveau α

Induktive Statistik

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

Nach der einfachen Testsituation wird jetzt das allgemeine Testproblem behandelt:

$$\Gamma = \Gamma_0 \cup \Gamma_1, \Gamma_0 \neq \emptyset \neq \Gamma_1, \Gamma_0 \cap \Gamma_1 = \emptyset.$$

mit den Hypothesen:

$$H_0 : \gamma \in \Gamma_0 \quad \text{und} \quad H_1 : \gamma \in \Gamma_1$$

Beispiele:

$$\begin{array}{ll} \mu \leq \mu_0 & \mu > \mu_0 \\ \mu = \mu_0 & \mu \neq \mu_0 \end{array}$$

Eine Entscheidungsfunktion δ ist **gleichmäßig bester Test zum Niveau α,**
wenn gilt

a) $P_I(\delta, \gamma_0) \leq \alpha$ für alle $\gamma_0 \in \Gamma_0$

b) für jeden anderen Test δ' vom Niveau α
$P_{II}(\delta, \gamma_1) \leq P_{II}(\delta', \gamma_1)$ für alle $\gamma_1 \in \Gamma_1$

Seien $\gamma_0 \in \Gamma_0$ und $\gamma_1 \in \Gamma_1$ beliebig ausgewählt, dann sind wegen a) δ und δ' auch Tests zum Niveau α für

$$(*) \quad H_0 : \gamma = \gamma_0 \text{ gegen } H_1 : \gamma = \gamma_1.$$

Daraus ergibt sich als **Ansatz** für die Bestimmung eines gleichmäßig
besten Tests zum Niveau α für das Ausgangsproblem:

1. Zu jedem Parameterpaar (γ_0, γ_1) mit $\gamma_0 \in \Gamma_0$ und $\gamma_1 \in \Gamma_1$ bestimmen wir den Neyman-Pearson-Test für das Problem $(*)$ mit einer für dieses Paar bzgl. α geeigneten Schranke für den Wahrscheinlichkeitsquotienten

$$\delta_{k_{\gamma_0,\gamma_1}}(x) = \begin{cases} d_0 : \gamma = \gamma_0 & Q_{\gamma_0,\gamma_1}(x) \leq k_{\gamma_0,\gamma_1} \\ d_1 : \gamma = \gamma_1 & Q_{\gamma_0,\gamma_1}(x) > k_{\gamma_0,\gamma_1} \end{cases}$$

mit

$$Q_{\gamma_0,\gamma_1}(x) = \frac{P_{\gamma_1}(X = x)}{P_{\gamma_0}(X = x)} \quad \text{bzw.} \quad Q_{\gamma_0,\gamma_1}(x) = \frac{f_{X,\gamma_1}(x)}{f_{X,\gamma_0}(x)}$$

2. Synthese dieser Tests zu einem Test:
   Da die Fehlerwahrscheinlichkeit 1. Art höchstens α sein soll, wird die Entscheidung d_1 nur getroffen, wenn man sich bei jedem Paar (γ_0, γ_1) für d_1 entscheidet. Also wird

$$\delta(x) = \begin{cases} d_0 & Q_{\gamma_0,\gamma_1}(x) \leq k_{\gamma_0,\gamma_1} \quad \text{für mindestens} \\ & \qquad\qquad\qquad\qquad \text{ein Paar } (\gamma_0, \gamma_1) \\ d_1 & Q_{\gamma_0,\gamma_1}(x) > k_{\gamma_0,\gamma_1} \quad \text{für alle Paare } (\gamma_0, \gamma_1) \end{cases}$$

   als Synthese-Test gewählt.

Beispiel: Binomialverteilung
$p = 0.1$: Ausschusswahrscheinlichkeit für ein hergestelltes Teil bei bisheriger Maschine
Behauptung des Anbieters einer neuen Anlage: $p < 0.1$

Hypothesen:

$$H_0 : p \geq \tilde{p} = 0.1 \qquad H_1 : p < \tilde{p} = 0.1$$

An fünf Wochentagen wird jeweils eine Stichprobe vom Umfang $m = 6$ entnommen und die Anzahl schlechter Teile festgestellt:

Die Anzahl schlechter Teile Y ist binomialverteilt mit Parameter p und $m = 6 \quad (B(6, p))$.
$n = 5$ Stichproben mit: $(x_1, x_2, x_3, x_4, x_5) = (0, 0, 0, 1, 0)$.
Für $p_1 < 0.1 \leq p_0$ ist der Wahrscheinlichkeitsquotient

$$\begin{aligned} Q(x) &= \frac{P_{p_1}(X = x)}{P_{p_0}(X = x)} = \frac{p_1^{\sum_{i=1}^n x_i} (1 - p_1)^{nm - \sum_{i=1}^n x_i}}{p_0^{\sum_{i=1}^n x_i} (1 - p_0)^{nm - \sum_{i=1}^n x_i}} \\ &= \left(\frac{p_1}{p_0}\right)^{\sum_{i=1}^n x_i} \left(\frac{1 - p_1}{1 - p_0}\right)^{nm - \sum_{i=1}^n x_i} \end{aligned}$$

monoton fallend $(p_1 < p_0)$ in $T(x) = \sum_{i=1}^n x_i$.

Also gilt:

$$Q_{p_0,p_1}(x) \leq k \Leftrightarrow \sum_{i=1}^{n} x_i \geq c_k.$$

für geeignetes c_k, das ganzzahlig gewählt werden kann, da $\sum_{i=1}^{n} x_i$ ganzzahlig ist.

Damit lässt sich der Neyman-Pearson-Test für $p = p_0$ gegen $p = p_1$ schrei-
ben als:

$$\delta_{p_0,p_1}(x) = \begin{cases} d_0 : p = p_0 & \sum_{i=1}^{n} x_i \geq c_{p_0,p_1} \\ d_1 : p = p_1 & \sum_{i=1}^{n} x_i < c_{p_0,p_1} \end{cases}$$

Das Niveau des Tests ist

$$P_I(\delta, p_0) = P_{p_0}(\sum_{i=1}^{n} X_i < c_{p_0,p_1})$$

Zu vorgegebenem α wird für jedes Paar (p_0, p_1)

$$c_{p_0,p_1}$$

so bestimmt, dass gilt:

$$P_I(\delta, p_0) = P_{p_0}(\sum_{i=1}^{n} X_i < c_{p_0,p_1}) \leq \alpha$$

(Wegen des Fehlers 2. Art wird die Schranke möglichst groß gewählt.)

Synthese der Tests:

$$\delta^*(x) = \begin{cases} d_0 : p \geq \tilde{p} & \displaystyle\sum_{i=1}^{n} x_i \geq c_{p_0,p_1} \\ & \text{für mindestens ein Paar } \ (p_0, p_1) \\ & \\ d_1 : p < \tilde{p} & \displaystyle\sum_{i=1}^{n} x_i < c_{p_0,p_1} \\ & \text{für alle Paare } \ (p_0, p_1) \end{cases}$$

...

...

oder mit $c := \min\{c_{p_0,p_1} \mid p_0 \geq \tilde{p}, p_1 < \tilde{p}\}$

$$\delta_c^*(x) = \begin{cases} d_0 : p \geq \tilde{p} & \sum\limits_{i=1}^{n} x_i \geq c \\ d_1 : p < \tilde{p} & \sum\limits_{i=1}^{n} x_i < c \end{cases}$$

Damit kann man auch die Tests δ_c^* für $c = 1, 2, \ldots$ vergleichen und ein geeignetes c auswählen, ohne zunächst zu jedem Paar (p_0, p_1) einen Test zu bestimmen.
Ein Vergleich basiert auf den Fehlerwahrscheinlichkeiten oder dem Verlauf von OC-Funktion bzw. Gütefunktion.

Die Fehlerwahrscheinlichkeit 1. Art wird maximal für $p = 0.1$, d.h.

$$\delta_c^* \text{ ist vom Niveau } \alpha_c = P_I(\delta_c^*, \tilde{p}) = G_{30,c-1}(0.1)$$

und natürlich jedem höheren Wert.
Für $c = 1$ ist $\alpha_c = 0.042...,$ für $c = 2$ erhalten wir $\alpha_c = 0.183...$
(siehe Abbildung)
Ist δ_c^* ein gleichmäßig bester Test zu diesem Niveau α_c?

Zu zeigen:
Ist δ ein weiterer Test zum Niveau α_c, so gilt

$$P_{II}(\delta_c^*, p_1) \leq P_{II}(\delta, p_1) \text{ für alle } p_1 < 0.1$$

δ_c^* und δ sind Tests zum Niveau α_c für

$(*) \qquad H_0 : p = 0.1 \quad$ gegen $\quad H_1 : p = p_1$, für jedes $p < 0.1$.

δ_c^* ist ein Neyman-Pearson-Test zum Niveau α_c, also ein bester Test zu
diesem Niveau.
Folglich gilt bezüglich der Testsituation $(*)$

$$P_{II}(\delta_c^*, p_1) \leq P_{II}(\delta, p_1)$$

Die Fehlerwahrscheinlichkeiten der Testsituation $(*)$ und der Ausgangstestsituation $H_0 : p \geq \tilde{p} = 0.1; H_1 : p < \tilde{p} = 0.1$ stimmen aber überein.
δ_c^* ist ein gleichmäßig bester Test zum Niveau α_c.

Entscheidend in diesem Beispiel ist:

1. Der Wahrscheinlichkeitsquotient ist für **jedes** Parameterpaar monoton fallend in der Statistik $T(x) = \sum x_i$.
2. Der Schwellenwert 0.1 gehört zu Γ_0.

Die Vorgehensweise führt auch zum Ziel, wenn der Wahrscheinlichkeitsquotient monoton steigend ist für jedes Parameterpaar (z.B. bei den Hypothesen $H_0 : p \leq 0.1$ gegen $H_1 : p > 0.1$). Wichtig ist die **Einheitlichkeit**.

Wir scheitern aber dann, wenn es ein Paar $(\gamma_0, \gamma_1), \gamma_0 \in \Gamma_0$
und $\gamma_1 \in \Gamma_1$, mit
monoton steigendem Wahrscheinlichkeitsquotienten und ein
Paar $(\gamma_0', \gamma_1'), \gamma_0' \in \Gamma_0$
und $\gamma_1' \in \Gamma_1$ mit monoton fallendem
Wahrscheinlichkeitsquotienten gibt, denn
die Tests

$$\delta_{\gamma_0,\gamma_1}(x) = \begin{cases} d_0 & \text{für} \quad T(x) \leq c \\ d_1 & \text{für} \quad T(x) > c \end{cases}, \text{ und } \delta_{\gamma_0',\gamma_1'}(x) = \begin{cases} d_0 & \text{für} \\ d_1 & \text{für} \end{cases}$$

...

...

führen zum Synthese-Test

$$\delta_{syn}(x) = \begin{cases} d_0 \text{ für } & T(x) \leq c \text{ oder } T(x) \geq c' \\ d_1 \text{ für } & c < T(x) < c' \end{cases}$$

δ_{syn} ist aber kein Neyman-Pearson-Test.

Vorgehensweise für die **Ermittlung eines gleichmäßigen besten Tests zum Niveau** α für ein Testproblem

$$H_0 : \gamma \in \Gamma_0 \quad \text{gegen} \quad H_1 : \gamma \in \Gamma_1$$

1. Suche eine suffiziente (und vollständige) Statistik T bezüglich γ.
2. Berechne für jedes Parameterpaar $(\gamma_0, \gamma_1), \gamma_0 \in \Gamma_0$ und $\gamma_1 \in \Gamma_1$, den Wahrscheinlichkeitsquotienten.
   Überprüfe das Monotonieverhalten des Wahrscheinlichkeitsquotienten in $T(x)$.

Falls

a) der Wahrscheinlichkeitsquotient **einheitlich monoton fallend** ist:

$$(1) \quad \delta_c^*(x) = \begin{cases} d_0 & T(x) \geq c \\ d_1 & T(x) < c \end{cases}$$

ist eine geeignete Entscheidungsfunktion.

b) der Wahrscheinlichkeitsquotient **einheitlich monoton steigend** ist:

$$(2) \quad \delta_c^*(x) = \begin{cases} d_0 & T(x) \leq c \\ d_1 & T(x) > c \end{cases}$$

ist eine geeignete Entscheidungsfunktion.

c) kein einheiliches Monotonieverhalten vorliegt:

Breche das Verfahren ab.

4. Bestimmung der Schranke c:
   Fehlerwahrscheinlichkeit 1. Art bei δ_c^* für einen Parameterwert γ_0 in Γ_0 : im Fall a)

$$P_I(\delta_c^*, \gamma_0) = P_{\gamma_0}(\delta_c^*(X) = d_1) = P_{\gamma_0}(T(X) < c)$$

im Fall b)

$$P_I(\delta_c^*, \gamma_0) = P_{\gamma_0}(\delta_c^*(X) = d_1) = P_{\gamma_0}(T(X) > c)$$

Suche Maximalstelle γ_0^* in Γ_0 und Maximalwert $\alpha_c^* = P_I(\delta_c^*, \gamma_0^*)$ für c.

Wähle c so, dass

- im stetigen Fall der Maximalwert mit dem gewünschtem Niveau übereinstimmt:
  $\alpha = \alpha_c^*$.
- im diskreten Fall der Maximalwert α_c^* dem gewünschten Niveau α am besten entspricht.

Ergebnis:
(1) bzw. (2) ist ein gleichmäßig bester Test zum Niveau $\alpha_c^* = P_I(\delta_c^*, \gamma_0^*)$.

Die Begründung entspricht genau der Argumentation im Beispiel.

Bemerkung:

Gibt es keine Maximalstelle und keinen Maximalwert für die Fehler-
wahrscheinlichkeit 1. Art, so gibt es ein Supremum, da die Fehlerwahrscheinlichkeit nach oben durch 1 beschränkt ist.
(1) bzw. (2) ist dann ein sinnvoller Test mit diesem Supremum
als Niveau, es kann aber nicht in der selben Weise gefolgert werden,
dass es sich um einen gleichmäßig besten Test zu diesem Niveau handelt.

Beispiel: Exponentialverteilung
Für eine Servicestelle soll getestet werden, ob die "Ankunftsrate" λ von Aufträgen größer als 5 ist.
Verteilungsannahme: Die Zeit zwischen den Ankunftszeitpunkten zweier Aufträge ("Zwischenankunftszeit") ist exponentialverteilt mit Parameter λ
Hypothesen:

$$H_0 : \lambda \leq 5 \quad \text{gegen} \quad H_1 : \lambda > 5$$

Entscheidungsgrundlage: Stichprobe mit Zurücklegen vom Umfang n.

1. Suffiziente und vollständige Statistik ist:

$$T(x) = \sum_{i=1}^{n} x_i$$

2. Für $\lambda_0 \leq 5$ und $\lambda_1 > 5$ ist

$$Q(x) = \frac{f_{X,\lambda_1}(x)}{f_{X,\lambda_0}(x)} = \frac{\prod\limits_{i=1}^{n} \lambda_1 \; e^{-\lambda_1 x_i}}{\prod\limits_{i=1}^{n} \lambda_0 \; e^{-\lambda_0 x_i}} = \left(\frac{\lambda_1}{\lambda_0}\right)^n e^{-\sum\limits_{i=1}^{n} x_i(\lambda_1 - \lambda_0)}$$

monoton fallend in $T(x) = \sum\limits_{i=1}^{n} x_i$.

3. Test

$$(1) \quad \delta_c^*(x) = \begin{cases} d_0 & T(x) = \sum\limits_{i=1}^{n} x_i \geq c \\ d_1 & T(x) = \sum\limits_{i=1}^{n} x_i < c \end{cases}$$

4.

$$\begin{aligned} P_I(\delta_c^*, \lambda_0) &= P_{\lambda_0}(\delta_c^*(X) = d_1) = P_{\lambda_0}\left(T(X) = \sum_{i=1}^{n} X_i < c \right) \\ &= F_{er(n),\lambda_0}(c) = 1 - e^{\,-\lambda_0 c} \sum_{i=1}^{n} \frac{(\lambda_0 c)^{i-1}}{(i-1)!} \end{aligned}$$

ist monoton steigend in λ_0.
Maximalstelle im Bereich $\leq 5 : \lambda^* = 5$
Maximalwert: $F_{er(n),5}(c)$

Ergebnis:

(1) ist gleichmäßig bester Test zum Niveau $F_{er(n),5}(c)$

α vorgegeben: $F_{er(n),5}(c) \stackrel{!}{=} \alpha$

c ist das $\alpha-$Quantil der Erlang-Verteilung mit Stufenzahl n und Parameter $\lambda = 5$

Für $n = 10$ und

$\alpha = 0.05 : \quad c = 1.085...$

In Kap. 9 wurde festgestellt, dass bei einigen Verteilungen das Monotonieverhalten des Wahrscheinlichkeitsquotienten von der Größenbeziehung der Parameter, genauer davon abhängt, ob

$$\gamma_0 < \gamma_1 \quad \text{oder} \quad \gamma_0 > \gamma_1$$

gilt.
Damit erhalten wir bei diesen Verteilungen ein einheitliches Monotonieverhalten des Wahrscheinlichkeitsquotienten für alle einseitigen Testprobleme:

(a) $H_0 : \gamma \leq \tilde{\gamma}$ gegen $H_1 : \gamma > \tilde{\gamma}$
bzw.
(b) $H_0 : \gamma \geq \tilde{\gamma}$ gegen $H_1 : \gamma < \tilde{\gamma}$

Bei (a) ist die Aufteilung des Parameterraums:

$$\Gamma_0 = (-\infty, \tilde{\gamma}], \quad \Gamma_1 = (\tilde{\gamma}, \infty)$$

bei (b)

$$\Gamma_0 = [\tilde{\gamma}, \infty), \quad \Gamma_1 = (-\infty, \tilde{\gamma})$$

Γ_1 liegt also jeweils nur auf *einer* Seite von Γ_0, daher die Bezeichnung
einseitig.

Es ist dann also bei der Ermittlung eines gleichmäßig besten Tests zu gegebenem Niveau Schritt 4 durchzuführen.

Bei den aufgeführten Beispielen ist die Maximalstelle jeweils der Schwellenwert $\tilde{\gamma}$.

Beispiel:

1. Poisson-Verteilung

$$H_0 : \lambda \leq \tilde{\lambda} \quad \text{gegen} \quad H_1 : \lambda > \tilde{\lambda}$$

Für $\lambda_0 < \lambda_1$ ist der Wahrscheinlichkeitsquotient monoton steigend in

$$T(x) = \sum_{i=1}^{n} x_i.$$

Der Test

$$(2) \qquad \delta_c^*(x) = \begin{cases} d_0 & T(x) = \sum\limits_{i=1}^{n} x_i \leq c \\ d_1 & T(x) = \sum\limits_{i=1}^{n} x_i > c \end{cases}$$

mit $c = 0, 1, 2, \ldots$ ist also Kandidat für einen gleichmäßig besten Test zu
einem noch zu bestimmenden, von c abhängigen Niveau.

Die Fehlerwahrscheinlichkeit 1. Art stimmt mit der Gütefunktion im Bereich Γ_0 überein.
Gütefunktion von δ_c^* :

$$
\begin{aligned}
G_{\delta_c^*}(\lambda) &= P_\lambda(\delta_c^*(X) = d_1) = P_\lambda(\sum_{i=1}^{n} X_i > c) \\
&= \sum_{k=c+1}^{\infty} P_\lambda(\sum_{i=1}^{n} X_i = k) = \sum_{k=c+1}^{\infty} \frac{(n\lambda)^k}{k!} e^{\,-n\lambda} \\
&= 1 - \sum_{k=0}^{c} \frac{(n\tilde{\lambda})^k}{k!} e^{\,-n\lambda}
\end{aligned}
$$

$$\max_{\lambda \leq \tilde{\lambda}} G_{\delta_c^*}(\lambda) = G_{\delta_c^*}(\tilde{\lambda}) \stackrel{!}{=} \alpha$$

$$\Rightarrow 1 - \sum_{k=0}^{c} \frac{(n\tilde{\lambda})^k}{k!} e^{-n\tilde{\lambda}} = \alpha$$

$$\sum_{k=0}^{c} \frac{(n\tilde{\lambda})^k}{k!} e^{-n\tilde{\lambda}} = 1 - \alpha$$

gesucht ist $(1-\alpha)$-Quantil der $Poi(n\tilde{\lambda})$-Verteilung

2. Normalverteilung

$$H_0 : \mu \geq \mu_0 \quad \text{gegen} \quad H_1 : \mu < \mu_0$$

Der Wahrscheinlichkeitsquotient ist einheitlich monoton fallend in $\sum_{i=1}^{n} x_i$.

$$(1) \qquad \delta^*_{c'}(x) = \begin{cases} d_0 & T(x) = \sum_{i=1}^{n} x_i \geq c' \\ d_1 & T(x) = \sum_{i=1}^{n} x_i < c' \end{cases}$$

ist Kandidat für einen gleichmäßig besten Test zu gegebenem Niveau bei
geeignet gewähltem c'.

Häufig formuliert man diesen Test mit dem Stichprobenmittelwert

$$(1) \qquad \delta_c^*(x) = \begin{cases} d_0 & \overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i \geq c \\ d_1 & \overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i < c \end{cases}$$

Gütefunktion:

$$\begin{aligned} G_{\delta_c*}(\mu) &= P_\mu(\delta_c^*(X) = d_1) = P_\mu\left(\frac{1}{n}\sum_{i=1}^{n} X_i < c\right) \\ &= P_\mu\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{c - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi(\sqrt{n}\frac{c - \mu}{\sigma}) \end{aligned}$$

Φ ist als Verteilungsfunktion monoton steigend, also ist die Gütefunktion
monoton fallend in μ.
Maximalstelle: μ_0
Maximalwert:

$$\alpha_c = P_I(\overline{X} < c) = \Phi(\sqrt{n}\frac{c - \mu_0}{\sigma})$$

δ_c^* ist also ein gleichmäßig bester Test zum Niveau α_c.
Zu einem vorgegebenem Niveau α kann jetzt c so bestimmt werden, dass
$\alpha_c = \alpha$ ist:

$$\alpha_c = \Phi\left(\sqrt{n}\frac{c-\mu_0}{\sigma}\right) = \alpha \iff \sqrt{n}\frac{c-\mu_0}{\sigma} = \Phi^{-1}(\alpha)$$

$$\iff c = \mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)$$

Zahlenbeispiel:

$$\mu_0 = 10, \sigma = 0.02, n = 10, \alpha = 0.05$$

$$u_{0.05} = -u_{0.95} = -1.645$$

$$c = 10 - \tfrac{0.02}{\sqrt{10}} 1.645 = 9.990$$