Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)

Lehrstuhl für Ökonometrie und Statistik

Kapitel XI - Operationscharakteristik und Gütefunktion

Induktive Statistik

Prof. Dr. W.-D. Heller

Hartwig Senska

Carlo Siebenschuh

Das Niveau eines Tests δ (Maximum der Fehlerwahrscheinlichkeit 1. Art) ist

$$\max_{\gamma \in \Gamma_0} P_\gamma(\delta(X) = d_1)$$

Zur Beurteilung eines Tests ist aber nicht nur dieser eine Wert interessant, sondern daneben ist natürlich auch die Fehlerwahrscheinlichkeit 1. Art für alle Parameterwerte $\gamma \in \Gamma_0$ und der Verlauf der Fehlerwahrscheinlichkeit 2. Art für $\gamma \in \Gamma_1$ von Bedeutung.

Beispiel:

Testproblem $H_0 : \lambda \geq \lambda_0$ gegen $H_1 : \lambda < \lambda_0$, Sei λ Parameter einer Exponentialverteilung, Gleichmäßig bester Test zum Niveau $\alpha = P_{\lambda_0}(\sum_{i=1}^{n} X_i > c)$ ist

$$\delta(x) = \begin{cases} d_0 & \sum_{i=1}^{n} x_i \leq c \\ d_1 & \sum_{i=1}^{n} x_i > c \end{cases}$$

Zu vorgegebenem α erhält man c mittels der Erlang-Verteilung.

Beispiel:

Parameter λ einer Exponentialverteilung

Für $\lambda < \lambda_0$ ist dann die Wahrscheinlichkeit für den Fehler 2. Art

$$\begin{aligned} P_\lambda(\sum_{i=1}^{n} X_i \leq c) &= F_{\sum\limits_{i=1}^{n} X_{i,\lambda}}(c) \\ &= 1 - e^{-\lambda c} \sum_{i=1}^{n} \frac{(\lambda c)^{i-1}}{(i-1)!}. \end{aligned}$$

(Für $\lambda \geq \lambda_0$ ist
$P_\lambda(\sum\limits_{i=1}^{n} X_i > c) = 1 - F_{\sum\limits_{i=1}^{n} X_{i,\lambda}}(c) = e^{-\lambda c} \sum\limits_{i=1}^{n} \frac{(\lambda c)^{i-1}}{(i-1)!}$

die Fehlerwahrscheinlichkeit 1.Art.)

Beispiel:

Parameter λ einer Exponentialverteilung

Die folgenden Tabellen geben für $c = 10$ und $c = 20$ die Werte dieser Wahrscheinlichkeit für verschiedene Werte von λ an.
Im Bereich $\lambda < \lambda_0$ handelt es sich um die Wahrscheinlichkeit für den Fehler 2. Art, für $\lambda \geq \lambda_0$ um die Wahrscheinlichkeit, die Nullhypothese richtigerweise nicht abzulehnen, je nach dem konkreten Wert von λ_0.
Betrachtet man diesen Wahrscheinlichkeitswert in Abhängigkeit von λ, so nennt man diese Funktion Operationscharakteristik oder OC-Funktion.

Beispiel: Wahrscheinlichkeiten für die Annahme der Nullhypothese

$c = 10$

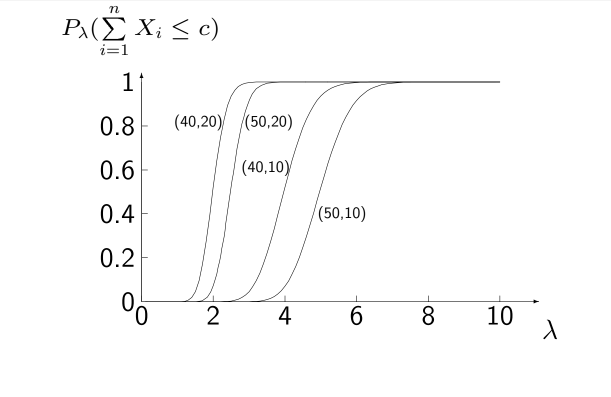
	$n =$					
λ	10	20	30	40	50	100
0.1	0.0	0.0	0.0	0.0	0.0	0.0
1	0.5421	0.0034	0.0	0.0	0.0	0.0
2	0.9950	0.5297	0.0218	0.00005	0.0	0.0
3	1	0.9781	0.5243	0.0462	0.00005	0.0
4	1	0.9998	0.9568	0.5210	0.0703	0.0
5	1	1	0.9991	0.9354	0.5188	0.0
6	1	1	1	0.9975	0.9156	0.0
7	1	1	1	1	0.9949	0.0004
8	1	1	1	1	0.9999	0.0171
9	1	1	1	1	1	1
10	1	1	1	1	1	1

Beispiel: Wahrscheinlichkeiten für die Annahme der Nullhypothese

$c = 20$

	$n =$					
λ	10	20	30	40	50	100
0.1	0.00005	0.0	0.0	0.0	0.0	0.0
1	0.9950	0.5297	0.0218	0.00005	0.0	0.0
2	1	0.9998	0.9568	0.5210	0.0703	0.0
3	1	1	1	0.9975	0.9156	0.0
4	1	1	1	1	0.9999	0.0171
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	1	1	1	1	1	1
9	1	1	1	1	1	1
10	1	1	1	1	1	1

Beispiel: OC-Funktionen für n=40, 50 und c=10, 20



Beispiel:

Parameter λ einer Exponentialverteilung

Man sieht in der Abbildung, dass für $\lambda_0 = 2$ beispielsweise die Werte

$n = 40, c = 20$ und $n = 50, c = 20$ in Frage kommen, wobei die Wahrscheinlichkeit für den Fehler 1.Art (also im Bereich $\lambda \geq 2$) bei $n = 40$ geringer ist als bei $n = 50$.
Analog für $\lambda_0 = 4$ mit $n = 50$ bzw. $n = 40$ und $c = 10$, wobei hier bei beiden Tests die Annahmewahrscheinlichkeit langsamer ansteigt. Also haben beide Tests eine geringere Trennschärfe.

Definition:

Sei δ eine Entscheidungsfunktion zu einem Testproblem mit Parameterraum Γ, so heißt die Funktion $L : \Gamma \to [0, 1]$ mit

$$L(\gamma) = P_{\gamma}(\delta(X) = d_0)$$

Operations-Charakteristik oder OC-Funktion.
Die OC-Funktion gibt die Wahrscheinlichkeit an, die Nullhypothese nicht abzulehnen, in Abhängigkeit vom Parameter γ

Beispiel:

Parameter p einer Bernoulli-Verteilung

Gleichmäßig bester Test zum Testproblem

$H_0 : p \leq p_0$ gegen $H_1 : p > p_0$

$$\delta(x) = \begin{cases} d_0 & \sum_{i=1}^{n} x_i \leq c \\ d_1 & \sum_{i=1}^{n} x_i > c \end{cases}$$

zum Niveau $P_{p_0}(\sum_{i=1}^{n} X_i > c)$.

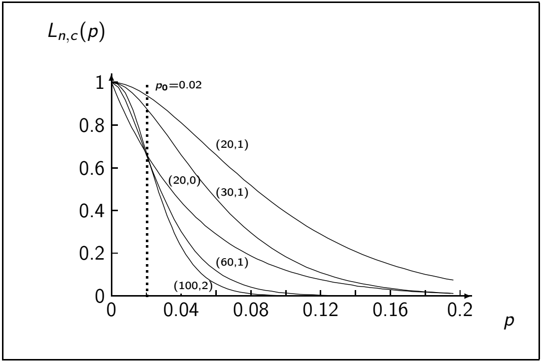
Beispiel:

Parameter p einer Bernoulli-Verteilung

Variieren kann man die Testschranke c und den Stichprobenumfang n.
Operations-Charakteristik zu diesem Test ist

$$L_{n,c}(p) = \sum_{k=0}^{c} \binom{n}{k} p^k (1-p)^{n-k}.$$

Die folgende Abbildung zeigt, daß die Nichtablehnwahrscheinlichkeit der Nullhypothese mit wachsendem p fällt. Soll auch beim Schwellenwert $p_0 = 0.02$ diese Wahrscheinlichkeit noch hoch sein, so ist die Fehlerwahrscheinlichkeit für ebenfalls hoch p knapp über 0.02 ebenfalls hoch. Ein schnelles Abfallen der OC-Funktion nach 0.02 kann durch geeignete Wahl von n und c erreicht werden.



Parameterwerte zum Beispiel

Die OC-Funktion liefert damit den Verlauf der Fehlerwahrscheinlichkeit 2. Art für $\gamma \in \Gamma_1$. Im Bereich $\gamma \in \Gamma_0$ gibt sie gerade die Annahmewahrscheinlichkeit bzw. Nichtablehnwahrscheinlichkeit der Nullhypothese an.

Da bei der Definition von Tests zum Niveau α der Fehler 1. Art im Vordergrund steht, betrachtet man häufig auch das wahrscheinlichkeitstheoretische Komplement

$$1 - L(\gamma).$$

- für $\gamma \in \Gamma_0$: Wahrscheinlichkeit für den Fehler 1. Art
- für $\gamma \in \Gamma_1$: Annahmewahrscheinlichkeit für die Gegenhypothese $\quad 1 - L(\gamma)$.

Bei Signifikanztests liegt das Augenmerk besonders auf der Gegenhypothese (d.h. Entscheidung für die Gegenhypothese soll mit hoher Sicherheit richtig sein), d.h. die Fehlerwahrscheinlichkeit 1. Art soll niedrig sein. Zudem sollte mit hoher Wahrscheinlichkeit die Gegenhypothese angenommen werden, wenn sie richtig ist (nachrangig gegenüber dem Fehler 1. Art).
Diese Eigenschaft (für $\gamma \in \Gamma_1$ mit hoher Wahrscheinlichkeit Entscheidung d_1 zu treffen) ist mit

$$G(\gamma) = 1 - L(\gamma)$$

zu überprüfen.

Definition:

Sei δ eine Entscheidungsfunktion zu einem Testproblem mit Parameterraum Γ, so heißt die Funktion
$G : \Gamma \to [0, 1]$ mit

$$G(\gamma) = P_\gamma(\delta(X) = d_1)$$

Gütefunktion (power) zu δ.

Beispiel:

Ein Hersteller von Distanzscheiben, die dazu dienen, das Wellenspiel in Getrieben auszugleichen, behauptet, dass seine Scheiben die Sollstärke von 5 mm mit einer Standardabweichung von $\sqrt{0.2}$ mm einhalten.
Wir nehmen an, dass die tatsächliche Stärke normalverteilt ist mit Mittelwert $\mu = 5$ mm und unbekanntem σ^2.
Zu prüfen ist, ob $\sigma^2 < 0.2$ statistisch gesichert ist.
Stichprobenumfang ist 30, Testniveau ist $\alpha = 0.05$.
Damit lautet das Testproblem:

$$H_0 : \sigma^2 \geq 0.2 = \tilde{\sigma}^2 \quad \text{gegen} \quad H_1 : \sigma^2 < 0.2$$

Beispiel:

Die Dichte der Normalverteilung ist:

$$f_{Y,\sigma^2}(y) = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y-\mu)^2},$$

also erhalten wir bei bekanntem μ eine einparametrige Exponentialfamilie mit

$$a(\sigma^2) = \frac{1}{\sigma}, h(y) = \frac{1}{\sqrt{2\pi}}, b(\sigma^2) = -\frac{1}{2\sigma^2}, \tau(y) = (y-\mu)^2$$

($b(\sigma^2)$ monoton steigend in σ^2)

Beispiel:

Damit ist

$$T(x) = \sum_{i=1}^{n} (x_i - \mu)^2$$

eine geeignete Testgröße

$$\delta(x) = \begin{cases} d_0 & \sum\limits_{i=1}^{n} (x_i - \mu)^2 > c \\ d_1 & \sum\limits_{i=1}^{n} (x_i - \mu)^2 \leq c \end{cases}$$

ist gleichmäßig bester Test zum Niveau

$$P_{\tilde{\sigma}^2=0.2} \left(\sum_{i=1}^{n} (X_i - \mu)^2 \leq c \right)$$

(Der Maximalwert der Fehlerwahrscheinlichkeit 1. Art liegt bei $\tilde{\sigma}^2 = 0.2$.)

Beispiel:

Da $\frac{X_i-\mu}{\sigma}$ $N(0,1)$-verteilt ist, folgt

$$\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 \sim \chi^2(n)\text{-verteilt.}$$

Damit gilt

$$
\begin{aligned}
& P_{\tilde{\sigma}^2=0.2}\left(\sum_{i=1}^{n}(X_i-\mu)^2 \leq c\right) \\
= \quad & P_{\tilde{\sigma}^2=0.2}\left(\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\tilde{\sigma}^2} \leq \frac{c}{\tilde{\sigma}^2}\right) = F_{\chi^2(n)}\left(\frac{c}{\tilde{\sigma}^2}\right) = \alpha = 0.05.
\end{aligned}
$$

Aus einer Tabelle für die χ^2-Verteilung:

$$F_{\chi^2(30)}(18.49) = 0.05.$$

Beispiel:

Folglich ist

$$c = \tilde{\sigma}^2 \cdot 18.49 = 0.2 \cdot 18.49 = 3.698$$

und

$$\delta^*(x) = \begin{cases} d_0 & \sum_{i=1}^{n} (x_i - \mu)^2 > 3.698 \\ d_1 & \sum_{i=1}^{n} (x_i - \mu)^2 \leq 3.698. \end{cases}$$

gleichmäßig bester Test zum Niveau $\alpha = 0.05$.

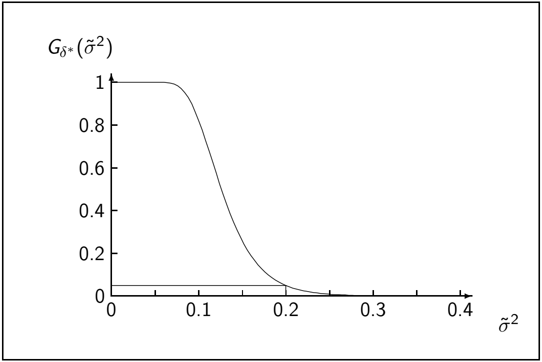
Dieser Test wird üblicherweise als "χ^2-Test" bezeichnet.

Beispiel:

Geänderte Testgröße: Mittlere quadratische Abweichung
$\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\mu)^2$.
Somit ist Testschranke dann:

$$\frac{1}{30}\cdot 3.698 = 0.123$$

Die quadratische Abweichung darf im Mittel nicht größer als 0.123 sein.
Die Gütefunktion zu diesem Test ist in der folgenden Abbildung dargestellt.
Wahrscheinlichkeit für den Fehler 1. Art für $\sigma^2 > 0.2$ sinkt sehr schnell ab.
Für $\sigma^2 < 0.2$ steigt die Wahrscheinlichkeit, die Gegenhypothese – richtigerweise – anzunehmen, für fallende Werte von σ^2 an.
Sie liegt aber z.B. für $\sigma^2 = 0.1$ immer noch bei nur etwa 0.8.



Parameterwerte zum Beispiel

Gütefunktion des Tests δ^*

Es wurde gezeigt: Bei einseitigen Testproblemen, (mit Statistik mit einheitlich monoton wachsendem Wahrscheinlichkeitsquotienten) steigt/fällt Wahrscheinlichkeit für Fehler 1. Art monoton im Parameter.

Daraus ergibt sich das Niveau α des Tests als Fehlerwahrscheinlichkeit 1. Art im Trennwert γ_0 desTestproblems.

Bei stetiger Gütefunktion folgt daraus, dass $1 - \alpha$ Supremum der Fehlerwahrscheinlichkeit zweiter Art ist.

Mehr ist nicht zu erreichen, wenn man sich auf das Niveau α festgelegt hat.

Also nur noch möglich, dass die Gütefunktion nach dem Trennwert schnell ansteigt.

Ein gleichmäßig bester Test δ^* zum Niveau α für $H_0 : \gamma \leq \gamma_0$ gegen $H_1 : \gamma > \gamma_0$ hat damit die Eigenschaft:

Die Gütefunktion jedes anderen Tests δ zum Niveau α liegt für $\gamma > \gamma_0$

nirgendwo oberhalb der von δ^*.

Mit Hilfe der Gütefunktion können Schwierigkeiten deutlich gemacht werden, die bei Suche nach gleichmäßig besten Tests zum Niveau α bei zweiseitigen Testproblemen entstehen.

Dies ist in folgendem Beispiel an einem zweiseitigen Testproblem zur Varianz einer Normalverteilung bei bekanntem Mittelwert verdeutlicht.

Beispiel:

Für die Normalverteilung bei bekanntem Mittelwert μ liege die folgende Testsituation vor:

$$H_0 : \sigma^2 = \tilde{\sigma}^2 = 0.2 \qquad \text{gegen} \qquad H_1 : \sigma^2 \neq \tilde{\sigma}^2 = 0.2 \ (*)$$

Dieser zweiseitige Test lässt sich zerlegen in zwei einseitige Tests:

$$H_0^{(1)} : \sigma^2 = \tilde{\sigma}^2 \qquad H_1^{(1)} : \sigma^2 > \tilde{\sigma}^2$$

und

$$H_0^{(2)} : \sigma^2 = \tilde{\sigma}^2 \qquad H_1^{(2)} : \sigma^2 < \tilde{\sigma}^2.$$

Beispiel:

Zu diesen Testproblemen erhalten wir gleichmässig beste
Tests zum Niveau
$\alpha > 0$ von folgender Art:
zu $H_0^{(1)}$ und $H_1^{(1)}$

$$\delta_1(x) = \begin{cases} d_1 & \sum_{i=1}^{n} (x_i - \mu)^2 > c_{1,\alpha} \\ d_0 & \sum_{i=1}^{n} (x_i - \mu)^2 \leq c_{1,\alpha}, \end{cases}$$

und zu $H_0^{(2)}$ und $H_1^{(2)}$

$$\delta_2(x) = \begin{cases} d_1 & \sum\limits_{i=1}^{n}(x_i - \mu)^2 \leq c_{2,\alpha} \\ d_0 & \sum\limits_{i=1}^{n}(x_i - \mu)^2 > c_{2,\alpha}. \end{cases}$$

Beispiel:

Angenommen: δ^* ist gleichmäßig bester Test zum Niveau α für die Ausgangssituation. Dann ist δ^* auch gleichmäßig bester Test zum Niveau α für die beiden einseitigen Tests.

Die Gütefunktion von δ^* müsste damit einen Verlauf nehmen, der für $\sigma^2 > 0.2$ mit der Gütefunktion von δ_1 und für $\sigma^2 < 0.2$ mit der von δ_2 übereinstimmt.

Da gleichmäßig beste Tests fast sicher für alle $\gamma \in \Gamma_1$ ($\sigma^2 > 0.2$ für δ_1 und alle $\sigma^2 < 0.2$ für δ_2) übereinstimmen folgt: δ^* stimmt bei Parameterwerten $\sigma^2 < 0.2$ mit δ_1 und bei Parameterwerten $\sigma^2 > 0.2$ mit δ_2 fast sicher überein.

Für $\alpha > 0$ ist aber $0 < c_{1,\alpha}, c_{2,\alpha} < \infty$ und die Tests δ_1 und δ_2 sind damit offensichtlich verschieden. Es kann also keinen Test geben, der sowohl mit δ_1, als auch mit δ_2 übereinstimmt.

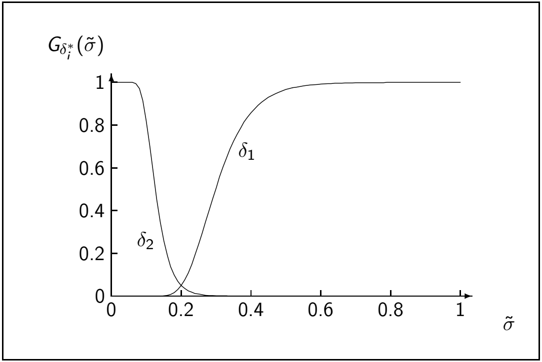
Folgerung:

Zu $0 < \alpha < 1$ gibt es keinen gleichmäßig besten Test Niveau α zum Testproblem

$$H_0 : \sigma^2 = \widehat{\sigma}_0^2 \quad \text{gegen} \quad H_1 : \sigma^2 \neq \widehat{\sigma}_0^2$$

Beispiel:
Die folgende Abbildung gibt Gütefunktionen der beiden einseitigen Tests wieder.
Man sieht, dass δ_1 als Entscheidungsfunktion zur Testsituation(*) zwar für $\sigma^2 > 0.2$ einen guten Verlauf der Gütefunktion hat, aber für $\sigma^2 < 0.2$ unbefriedigend,
da die Wahrscheinlichkeit die Nullhypothese abzulehnen, klein ist und schnell gegen Null geht.
Umgekehrt ist δ_2 als Test für (*) gut im Bereich $\sigma^2 < 0.2$, aber schlecht im Bereich $\sigma^2 > 0.2$.

Beispiel:



Parameterwerte zum Beispiel

Gütefunktion der Tests δ_1 und δ_2