# Verteilungsfreie Verfahren, SS 2004

**Leo Knüsel, Universität München**

# Inhaltsverzeichnis


1. Asymptotische relative Effizienz zwischen zwei Schätzfunktionen ..... 1
   1. Problemstellung ..... 1
   2. Varianz des Zentralwerts ..... 1
   3. Asymptotische Verteilung des Zentralwerts, Teil 1 ..... 4
   4. Asymptotische Verteilung des Zentralwerts, Teil 2 ..... 6
   5. Der Spezialfall mit $f(\tilde{\mu}) = 0$ ..... 7
   6. Asymptotische Verteilung von beliebigen Quantilen ..... 8
   7. Relative Effizienz von Schätzfunktionen ..... 9
   8. Asymptotische relative Effizienz von Schätzfunktionen ..... 10
   9. Beispiele ..... 13

2. Asymptotische relative Effizienz zwischen zwei Testverfahren ..... 14
   1. Problemstellung ..... 14
   2. Asymptotische Gütefunktion des t-Tests ..... 15
   3. Asymptotische Gütefunktion des Vorzeichentests ..... 15
   4. Asymptotische relative Effizienz ..... 16
   5. Vergleich des t-Tests mit dem Zeichentest bei endlichem Stichprobenumfang ..... 19

3. Median-Verfahren ..... 24
   1. Allgemeine Definition des Medians ..... 24
   2. Vertrauensgrenzen für den Median ..... 24
   3. Vertrauensgrenzen für beliebige Quantile ..... 26
   4. Randomisierte Vertrauensgrenzen ..... 27
   5. Ein-Stichproben-Problem für den Median ..... 29
   6. Zwei-Stichproben-Problem für den Median ..... 30
   7. Simulationsexperiment zur Lebensdauer eines Gerätes ..... 31

4. Vorzeichentest ..... 34
   1. Vorzeichentest ..... 34
   2. Asymptotische Gütefunktion des Tests A ..... 35
   3. Asymptotische Gütefunktion des Tests B ..... 36
   4. Vergleich der asymptotischen Gütefunktionen ..... 38
   5. Vergleich bei endlichem Stichprobenumfang ..... 39

5. Test von McNemar ..... 43
   1. Problemstellung, Variante 1 ..... 43
   2. Problemstellung, Variante 2 ..... 46
   3. Eigenschaften des Tests von McNemar ..... 47

6. Unabhängige oder verbundene Stichproben? Vierfeldertest oder Test von McNemar? ... 49
   1. Problemstellung ..... 49
   2. Versuchsplan mit unabhängigen Stichproben; Vierfeldertest ..... 49
   3. Versuchsplan mit verbundenen Stichproben; Test von McNemar ..... 50
   4. Asymptotische relative Effizienz ..... 52
   5. Vergleich bei endlichem Stichprobenumfang ..... 54

7. Lineare Rangtests beim Zweistichprobenproblem ..... 58
1. Ordnungsstatistiken und Rangstatistiken (order and rank statistics) ..... 58
2. Klassischer doppelter t-Test ..... 59
3. Lineare Rangtests beim Lagemodell ..... 60
4. Asymptotische Gütefunktion beim t-Test ..... 62
5. Asymptotische Gütefunktion beim allgemeinen linearen Rangtest ..... 63
6. Asymptotische relative Effizienz, Beispiele ..... 66
7. Lokal optimale Rangtests (locally most powerful rank tests) ..... 72
8. Permutationstest ..... 75

8. Test von Wilcoxon-Mann-Whithey bei beliebigen Verteilungen ..... 76
1. Problemstellung, stochastische Ordnung ..... 76
2. Test von Wilcoxon-Mann-Whitney im stetigen Fall ..... 78
3. Test von Wilcoxon-Mann-Whitney im allgemeinen Fall ..... 79

9. Lineare Rangtests beim Einstichprobenproblem ..... 83
1. Problemstellung ..... 83
2. Allgemeine lineare Rangtests ..... 85
3. Andere Darstellung von $L_+$ beim Test von Wilcoxon ..... 86
4. Asymptotische Gütefunktion beim einfachen t-Test ..... 87
5. Asymptotische Gütefunktion bei allgemeinen linearen Rangtests ..... 87
6. Asymptotische relative Effizienz, Beispiele ..... 90

Anhang ..... 94
A1 Hilfssatz zur stochastischen Konvergenz ..... 94
A2 Vier-Felder-Test, Asymptotische Gütefunktion ..... 94
A3 Zur asymptotischen Gütefunktion beim Zweistichprobenproblem ..... 98
A4 Optimale Gewichtsfunktion ..... 100
A5 Zum Test von Wilcoxon ..... 101
A6 Zum Normal-Scores-Test ..... 103
A7 Zur Gleichverteilung ..... 105
A8 Zur stochastischen Ordnung ..... 106
A9 Zum Test von Wilcoxon-Mann-Whitney ..... 109
A10 Zur asymptotischen Gütefunktion beim Einstichprobenproblem ..... 111
A11 Optimale Gewichtsfunktion beim Einstichprobenproblem ..... 113

Literatur ..... 114

# 1. Asymptotische relative Effizienz zwischen zwei Schätzfunktionen

## 1. Problemstellung

Es sei

$X$ eine stetige Zufallsgröße mit Verteilungsfunktion $F$ und Dichte $f$;

Erwartungswert von $X$: $\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$;

Median von $X$: $\tilde{\mu} = med(X)$: $F(\tilde{\mu}) = \frac{1}{2}$.

Wir wollen annehmen, dass der Erwartungswert und der Median übereinstimmen:
$\mu = E(X) = med(X)$; dies ist z.B. der Fall, wenn die Dichte symmetrisch ist bezüglich $\mu$.

*Ziel:* $\mu$ schätzen aufgrund einer unabhängigen Stichprobe vom Umfang $n$.

Es seien also

$X_1, \cdots, X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion $F$;

$\mu = E(X_i) = med(X_i)$ und $\sigma^2 = Var(X_i)$, $0 < \sigma^2 < \infty$;

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (arithmetisches Mittel);

$\tilde{X} = \text{Ze}(X_1, \cdots, X_n)$ (Zentralwert).

Dann gilt:

$E(\bar{X}) = \mu$ und $Var(\bar{X}) = \sigma^2 / n$

$E(\tilde{X}) = \mu$ und $Var(\tilde{X}) = ?$

$\bar{X}$ und $\tilde{X}$ sind beide erwartungstreue Schätzfunktionen für $\mu$.

$\bar{X}$ ist besser als $\tilde{X}$, wenn $\text{Var}(\bar{X}) < \text{Var}(\tilde{X})$.

Die Varianz von $\bar{X}$ und von $\tilde{X}$ hängt ab von der Verteilungsfunktion $F$.

Für welche Verteilungsfunktionen $F$ ist $\bar{X}$ besser als $\tilde{X}$?

Um diese Frage beantworten zu können, brauchen wir die Varianz des Zentralwerts $\tilde{X}$.

## 2. Varianz des Zentralwerts

Es seien

$X_1, \cdots, X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion $F$ und der Dichte $f$;

$$\tilde{X} = Ze(X_1, \ldots, X_n) = \begin{cases} X_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}\left( X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)} \right) & \text{falls } n \text{ gerade.} \end{cases}$$

Wir wollen die Varianz von $\tilde{X}$ berechnen, und dazu benötigen wir die Dichte von $\tilde{X}$.

Verteilungsfunktion von $\tilde{X}$:

$G(x) = P\{\tilde{X} < x\} = ?$

Wir wollen annehmen, dass $n$ ungerade ist, d.h. $n = 2k+1$. Nun sei

$N_x = \#\{i \,|\, X_i < x\}$.

Dann gilt

$N_x \sim Bi(n, p)$ mit $p = F(x) = P\{X_i < x\}$;

$\{\tilde{X} < x\} = \left\{ X_{(\frac{n+1}{2})} < x \right\} = \{X_{(k+1)} < x\} = \{N_x \geq k+1\}$.

Somit gilt

$$G(x) = P\{\tilde{X} < x\} = P\{N_x \geq k+1\} = \sum_{j=k+1}^{n} \binom{n}{j} p^j (1-p)^{n-j} \quad \text{mit } p = F(x);$$

$$g(x) = \frac{d}{dx} G(x);$$

$$E(\tilde{X}) = \tilde{\mu} = \int_{-\infty}^{\infty} x\, g(x)\, dx;$$

$$Var(\tilde{X}) = \tau^2 = \int_{-\infty}^{\infty} x^2\, g(x)\, dx - \tilde{\mu}^2.$$

Die Integrale für Erwartungswert und Varianz lassen sich in vielen Fällen numerisch berechnen mit Hilfe von Maple. In Tabelle 1 finden sich die Ergebnisse für die folgenden fünf Verteilungen:

a) Normalverteilung $N(0,1)$: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $-\infty < x < \infty$;

b) Cauchyverteilung $C(0,1)$: $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$;

c) Logistische Verteilung: $f(x) = \frac{e^{-x}}{\left(1+e^{-x}\right)^2}$, $-\infty < x < \infty$;

d) Laplaceverteilung: $f(x) = \frac{1}{2} e^{-|x|}$, $-\infty < x < \infty$;

e) Rechtecksverteilung $R(-1,1)$: $f(x) = \frac{1}{2}$, $-1 < x < 1$.

**Tabelle 1:** Varianz des Zentralwerts; die Tabelle gibt die Werte $n\tau^2$, wobei $\tau^2 = Var(Ze)$

| $n$ | $N(0,1)$ | $C(0,1)$ | Laplace | logistisch | $R(-1,1)$ |
|---|---|---|---|---|---|
| 1 | 1 | $\infty$ | 2 | $\frac{1}{3}\pi^2 = 3.290$ | $\frac{1}{3} = 0.333$ |
| 3 | 1.346 | $\infty$ | 1.917 | 3.870 | $\frac{3}{5} = 0.6$ |
| 5 | 1.434 | 6.106 | 1.756 | 3.949 | 0.714 |
| 11 | 1.509 | 3.375 | 1.521 | 3.989 | 0.846 |
| 21 | 1.538 | 2.870 | 1.374 | 3.997 | 0.913 |
| 51 | 1.558 | 2.618 | 1.236 | 3.999 | 0.962 |
| 101 | 1.564 | 2.541 | 1.165 | 4.000 | 0.981 |
| $\infty$ | $\frac{1}{2}\pi = 1.571$ | $\frac{1}{4}\pi^2 = 2.467$ | 1 | 4 | 1 |

Man beachte, dass für $n = 1$ der Zentralwert mit $X_1$ übereinstimmt, und dass dann $\tau^2 = Var(Ze) = Var(X_1) = \sigma^2$ ist. Für $n > 1$ ist $\text{Var}(\bar{X}) < \text{Var}(\tilde{X})$, falls $\sigma^2 < n\tau^2$, und Tabelle 1 zeigt, dass dies der Fall ist bei der Normalverteilung, bei der Rechtecksverteilung und bei der logistischen Verteilung, jedoch nicht bei der Cauchy- und Laplaceverteilung.

Für gerade Werte von *n* wird die Berechnung der Varianz etwas komplizierter (vgl. Büning, 1994, Seite 58). Tabelle 1 zeigt, dass $n\tau^2$ für $n \to \infty$ gegen einen Grenzwert konvergiert. Diesen Grenzwert wollen wir in den folgenden Abschnitten bestimmen.

*Maple-Arbeitsblatt* `maple1.mws` *zur Berechnung der Werte in Tabelle 1*

```
> #
  # Varianz des Zentralwerts
  #
> restart;
> #
  # Normalverteilung und Fehlerfunktion
> Phi:=z->1/2+(1/2)*erf(z/sqrt(2));
> Phic:=z->(1/2)*erfc(z/sqrt(2));
> phi:=z->exp(-z*z/2)/sqrt(2*Pi);
> #
  # Normalverteilung
  F := x -> Phi(x);
> #
  # Logistische Verteilung
  # F := x -> exp(x)/(1.0+exp(x));
> #
  # Cauchy-Verteilung
  # f := x -> 1/(Pi*(1+x^2));
  # F := x -> int(f(t),t=-infinity..x);
> # F(x);
> # int(x*f(x),x=0..infinity);
> # plot(F(x),x=-5..5);
> #
> # Laplaceverteilung
> # f:=x->(1/2)*exp(-abs(x));
> # plot(f(x),x=-3..3);
> # F := x -> int(f(t),t=-infinity..x);
> # F(x);
> #
  # Rechtecksverteilung auf [-1,1]
  # f:=x->piecewise(x<-1,0,x>1,0,1/2);
> # f(x);
> # plot(f(x),x=-2..2);
> # F := x -> int(f(t),t=-1..x);
> # F(x);
> #
  # X1,...,Xn unabhängig, je mit Vert.funktion F;
  # Y = Ze(X1,...,Xn) = Zentralwert von X1,...,Xn
  # G := Vert.funktion von Y
  # Erwartungswert von Y:
  #   E(Y) = int(y*g(y),y=-infinity..infinity) = mu
  # Varianz von Y :
  #   Var(Y) = int((y-mu)^2*g(y),y=-infinity..infinity) = tau2
  #
> #
  # Verteilung des Zentralwerts Y = Ze(X1,...,Xn) = X(k+1) falls n=2k+1;
  # G := Vert.funktion von Y
> G := y -> sum(binomial(n,j)*F(y)^j*(1-F(y))^(n-j),j=k+1..n);
> g := y -> diff(G(y),y);
> g(y);
> k:=0;
> n:=2*k+1;
> # existiert der Erw.wert?
  int(abs(y)*g(y),y=-infinity..infinity);
  evalf(%);
> mu := int(y*g(y),y=-infinity..infinity);
> evalf(%);
> tau2 := int(y^2*g(y),y=-infinity..infinity);
> evalf(%);
> n*%;
> evalf(Pi/2);
>
```

## 3. Asymptotische Verteilung des Zentralwerts, Teil 1

Wir wollen wieder annehmen, dass $n = 2k+1$ (d.h. $n$ ist ungerade), und wir wollen die Verteilung des Zentralwerts $\tilde{X}$ für $n \to \infty$ bestimmen. Wir haben im Abschnitt 2 gesehen, dass

$$G_n(x) = P\{\tilde{X} < x\} = P\{N_x \geq k+1\}$$

wobei

$$N_x = \#\{i \,|\, X_i < x\} \sim Bi(n,p) \text{ mit } p = F(x).$$

Wir standardisieren die Zufallsgröße $N_x$ und erhalten

$$(1)\quad G_n(x) = P\{N_x \geq k+1\} = P\left\{\frac{N_x - np}{\sqrt{np(1-p)}} \geq \frac{k+1-np}{\sqrt{np(1-p)}}\right\}.$$

Nun gilt aufgrund des zentralen Grenzwertsatzes

$$(2)\quad \frac{N_x - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1),$$

und wir haben zu untersuchen, was mit dem Term

$$(*) = \frac{k+1-np}{\sqrt{np(1-p)}}$$

passiert beim Grenzübergang $n \to \infty$. Es ist $k+1 = \frac{n+1}{2}$, und daher gilt

$$(*) = \frac{\frac{n+1}{2} - np}{\sqrt{np(1-p)}} = \frac{n\left(\frac{1}{2} - p + \frac{1}{2n}\right)}{\sqrt{np(1-p)}} = \sqrt{n}\,\frac{\frac{1}{2} - p + \frac{1}{2n}}{\sqrt{p(1-p)}}$$

$$(3)\quad \xrightarrow[n\to\infty]{} \begin{cases} +\infty & \text{falls } p < \frac{1}{2} \text{ d.h. falls } F(x) < \frac{1}{2} \\ 0 & \text{falls } p = \frac{1}{2} \text{ d.h. falls } F(x) = \frac{1}{2} \\ -\infty & \text{falls } p > \frac{1}{2} \text{ d.h. falls } F(x) > \frac{1}{2}. \end{cases}$$

Wenn der Median $\tilde{\mu} = med(F)$ eindeutig bestimmt ist, so gilt

$$F(x) \begin{cases} < \frac{1}{2} & \text{falls } x < \tilde{\mu} \\ = \frac{1}{2} & \text{falls } x = \tilde{\mu} \\ > \frac{1}{2} & \text{falls } x > \tilde{\mu}, \end{cases}$$

und somit folgt aus (1), (2) und (3)

$$(4)\quad G_n(x) \xrightarrow[n\to\infty]{} \begin{cases} 0 & \text{für } x < \tilde{\mu} \\ \frac{1}{2} & \text{für } x = \tilde{\mu} \\ 1 & \text{für } x > \tilde{\mu}, \end{cases}$$

und damit gilt $\tilde{X} \xrightarrow[n\to\infty]{P} \tilde{\mu} = med(F)$. Wenn ein Intervall $[a,b]$ existiert mit

$$(5)\quad F(x) \begin{cases} < \frac{1}{2} & \text{falls } x < a \\ = \frac{1}{2} & \text{falls } a \leq x \leq b \\ > \frac{1}{2} & \text{falls } x > b, \end{cases}$$

so ist der Median von $F$ nicht eindeutig bestimmt; jeder Wert $\tilde{\mu} \in [a,b]$ ist dann ein Median von $F$. Dann folgt aus (1), (2) und (3)

$$(6)\quad G_n(x) = P\{\tilde{X} < x\} \xrightarrow[n\to\infty]{} G(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{1}{2} & \text{für } x \in [a,b] \\ 1 & \text{für } x > b. \end{cases}$$

Wir haben also hier das Resultat $\tilde{X} \xrightarrow{d} G$, wobei die Grenzverteilung $G$ der symmetrischen Zweipunktverteilung auf $\{a,b\}$ entspricht: $P\{Y = a\} = P\{Y = b\} = \frac{1}{2}$.

*Ergänzungen*

a) Wenn der Stichprobenumfang *n* gerade ist d.h. $n = 2k$ , dann gilt

$$\tilde{X} = \tfrac{1}{2}\left(X_{(k)} + X_{(k+1)}\right),$$

und wir haben

$$X_{(k)} \le \tilde{X} \le X_{(k+1)}.$$

Daher gilt

(7) $$\left\{X_{(k+1)} < x\right\} \subset \left\{\tilde{X} < x\right\} \subset \left\{X_{(k)} < x\right\}.$$

Wenn der Median von *F* nicht eindeutig bestimmt ist wie in (5), so gilt analog zu (6)

(8)
$$P\left\{X_{(k)} < x\right\} \xrightarrow[n\to\infty]{} G(x)$$
$$P\left\{X_{(k+1)} < x\right\} \xrightarrow[n\to\infty]{} G(x),$$

wobei

$$G(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{1}{2} & \text{für } x \in [a,b] \\ 1 & \text{für } x > b. \end{cases}$$

Somit folgt aus (7) und (8)

$$G_n(x) = P\left\{\tilde{X} < x\right\} \xrightarrow[n\to\infty]{} G(x).$$

Die Beziehung (6) gilt also auch für gerade Werte von *n*. Falls der Median $\tilde{\mu} = med(F)$ eindeutig bestimmt ist, so folgt $\tilde{X} \xrightarrow{P} \tilde{\mu}$ für $n \to \infty$.

b) Die Ergebnisse, die wir hier für den Zentralwert gefunden haben, gelten in analoger Weise auch für allgemeine Stichprobenquantile. Es sei *q* ein Quantum mit $0 < q < 1$, und das Verteilungsquantil $x_q$ sei so definiert, dass $F(x_q) = q$. Wenn ein Intervall $[a,b]$ existiert mit

(9) $$F(x) \begin{cases} < q & \text{falls } x < a \\ = q & \text{falls } a \le x \le b \\ > q & \text{falls } x > b, \end{cases}$$

so ist das Quantil $x_q$ von *F* nicht eindeutig bestimmt; jeder Wert $x_q \in [a,b]$ ist dann ein *q*-Quantil von *F*. Das Stichprobenquantil zum Quantum *q* wollen wir wie folgt definieren. Es sei

$n =$ Stichprobenumfang;
$q =$ vorgegebenes Quantum ( $0 < q < 1$);
$\tilde{k} = (n+1)q = k + \delta,$ wobei *k* ganzzahlig und $0 \le \delta < 1$.

Dann ist das Stichprobenquantil zum Quantum *q* definiert als

$$Q = X_{(k)} + \delta\left(X_{(k+1)} - X_{(k)}\right).$$

Für $q = \frac{1}{2}$ erhalten wir die klassische Definition des Zentralwerts. Nun gilt analog zu (6)

(10) $$G_n(x) = P\{Q < x\} \xrightarrow[n\to\infty]{} G(x) = \begin{cases} 0 & \text{für } x < a \\ q & \text{für } x \in [a,b] \\ 1 & \text{für } x > b. \end{cases}$$

Wir haben also hier das Resultat $Q \xrightarrow{d} G$ für $n \to \infty$, wobei die Grenzverteilung *G* einer Zweipunktverteilung auf $\{a,b\}$ entspricht: $P\{Y = a\} = q$, $P\{Y = b\} = 1 - q$. Falls das Quantil $x_q$ eindeutig definiert ist, so ist $a = b$ und wir erhalten das Ergebnis $Q \xrightarrow{P} x_q$ für $n \to \infty$.

## 4. Asymptotische Verteilung des Zentralwerts, Teil 2

Wie im Abschnitt 3 wollen wir annehmen, dass $X_1,\dots,X_n$ unabhängige Zufallsgrößen sind je mit der Verteilungsfunktion *F* und mit der Dichtefunktion *f*. Es sei $\mu$ der Erwartungswert, $\sigma^2$ die Varianz und $\tilde{\mu}$ der Median von *F*. Weiter sei

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \quad = \text{ arithmetisches Mittel von } X_1,\dots,X_n$$

$$\tilde{X} = Ze(X_1,\dots,X_n) \quad = \text{ Zentralwert von } X_1,\dots,X_n.$$

Aufgrund des schwachen Gesetzes der großen Zahlen gilt

(11) $\overline{X} \xrightarrow[n\to\infty]{P} \mu$

und im vorangehenden Abschnitt haben wir gesehen, dass auch

(12) $\tilde{X} \xrightarrow[n\to\infty]{P} \tilde{\mu}$,

falls der Median $\tilde{\mu}$ eindeutig definiert ist. Wenn nun Erwartungswert und Median übereinstimmen (d.h. wenn $\mu = \tilde{\mu}$), so konvergieren beide Schätzfunktionen $\overline{X}$ und $\tilde{X}$ in Wahrscheinlichkeit gegen den gleichen Parameter, und aufgrund der asymptotischen Ergebnisse (11) und (12) ist es nicht möglich zu entscheiden, ob das arithmetische Mittel besser ist als der Zentralwert. Die Grenzverteilung von $\overline{X}$ und $\tilde{X}$ ist ja in beiden Fällen die Einpunktverteilung bei $\mu$.

Um eine echte asymptotische Vergleichsmöglichkeit zu finden, wollen wir die Schätzfunktionen $\overline{X}$ und $\tilde{X}$ so skalieren, dass als Grenzverteilungen nicht-ausgeartete Verteilungen auftreten. Wir betrachten also die Zufallsgrößen $\sqrt{n}(\overline{X} - \mu)$ und $\sqrt{n}(\tilde{X} - \tilde{\mu})$. Aufgrund des zentralen Grenzwertsatzes gilt

$$\sqrt{n}\left(\overline{X} - \mu\right) \xrightarrow{d} N(0,\sigma^2),$$

und wir interessieren uns nun für die Grenzverteilung von $\sqrt{n}(\tilde{X} - \tilde{\mu})$. Es wird sich zeigen, dass diese Grenzverteilung im allgemeinen auch eine Normalverteilung ist, und der Vergleich der Varianzen liefert dann das gesuchte Gütekriterium.

Wir wollen wiederum annehmen, dass *n* ungerade sei, und weiter sei sei $Y_n = \sqrt{n}(\tilde{X} - \tilde{\mu})$. Dann gilt analog zu (1) und (3)

$$\begin{aligned} G_n(y) &= P\{Y_n < y\} = P\left\{\sqrt{n}\left(\tilde{X} - \tilde{\mu}\right) < y\right\} = P\left\{\tilde{X} < \tilde{\mu} + \frac{y}{\sqrt{n}}\right\} \\ &= P\left\{\tilde{X} < x\right\} \text{ mit } x = \tilde{\mu} + \frac{y}{\sqrt{n}} \\ &= P\left\{N_x \geq \tfrac{n+1}{2}\right\} \text{ wobei } N_x = \#\left\{i \,\middle|\, X_i < x\right\} \sim Bi(n,p) \text{ mit } p = F(x) \\ &= P\left\{\frac{N_x - np}{\sqrt{np(1-p)}} \geq (*)\right\} \text{ wobei } (*) = \sqrt{n}\,\frac{\frac{1}{2} - p + \frac{1}{2n}}{\sqrt{p(1-p)}}. \end{aligned} \tag{13}$$

Nun ist $F(\tilde{\mu}) = \frac{1}{2}$ und

$$p = F(x) = F\left(\tilde{\mu} + \frac{y}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{} F(\tilde{\mu}) = \tfrac{1}{2},$$

und wir erhalten

$$\sqrt{n}\left(p - \tfrac{1}{2}\right) = \sqrt{n}\left[F\left(\tilde{\mu} + \frac{y}{\sqrt{n}}\right) - F(\tilde{\mu})\right] = y\,\frac{F\left(\tilde{\mu} + \frac{y}{\sqrt{n}}\right) - F(\tilde{\mu})}{y/\sqrt{n}} \xrightarrow[n\to\infty]{} y\,f(\tilde{\mu}).$$

Somit gilt für $n \to \infty$

$$(*) = \sqrt{n}\,\frac{\frac{1}{2} - p + \frac{1}{2n}}{\sqrt{p(1-p)}} \quad \longrightarrow \quad -2y\,f(\tilde{\mu}),$$

und wir erhalten aus (13)

$$G_n(y) = P\left\{\frac{N_x - np}{\sqrt{np(1-p)}} \geq (*)\right\} \quad \xrightarrow[n\to\infty]{} \quad G(y) = P\left\{Z > -2y\,f(\tilde{\mu})\right\} \text{ wobei } Z \sim N(0,1).$$

Die Grenzverteilung *G* kann auch geschrieben werden als

$$G(y) \;=\; P\left\{Z < 2y\,f(\tilde{\mu})\right\} \;=\; P\left\{\frac{Z}{2f(\tilde{\mu})} < y\right\},$$

und da die Zufallsgröße $Z/[2f(\tilde{\mu})]$ eine Normalverteilung $N(0,\tau^2)$ besitzt mit $\tau^2 = 1/\left[4f^2(\tilde{\mu})\right]$, so können wir auch schreiben

(14) $Y_n = \sqrt{n}(\tilde{X} - \tilde{\mu}) \quad \xrightarrow[n\to\infty]{d} \quad N(0,\tau^2)$ mit $\tau^2 = \dfrac{1}{4f^2(\tilde{\mu})}$.

Wie im Abschnitt 3 kann man zeigen, dass dieses Ergebnis nicht nur für ungerade Werte von *n* gilt sondern auch für gerade.

## 5. Der Spezialfall mit $f(\tilde{\mu}) = 0$

Die Beziehung (14) ist nur sinnvoll, wenn $f(\tilde{\mu}) > 0$ ist, da ja $\tau^2 = \infty$ für $f(\tilde{\mu}) = 0$. Wir möchten jetzt den Spezialfall $f(\tilde{\mu}) = 0$ genauer untersuchen. Es seien

$X_1, \ldots, X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion *F* und der Dichte $f$;

$\tilde{\mu} = med(F)$, d.h. $F(\tilde{\mu}) = \frac{1}{2}$;

$f(\tilde{\mu}) = 0, \quad f'(\tilde{\mu}) = 0, \quad f''(\tilde{\mu}) > 0.$

Die Dichtefunktion *f* hat also beim Median $\tilde{\mu}$ ein Minimum mit dem Wert 0. Es sei $\tilde{X}$ der Zentralwert von $X_1, \ldots, X_n$, und wir nehmen wiederum der Einfachheit halber an, dass *n* ungerade sei. Wir schreiben $Y_n = n^{\alpha}\left(\tilde{X} - \tilde{\mu}\right)$, und wir wollen $\alpha$ so bestimmen, dass die Zufallsgröße $Y_n$ eine nichtausgeartete Grenzverteilung besitzt. Analog zu (13) gilt

$$\begin{aligned} G_n(y) &= P\{Y_n < y\} = P\left\{n^{\alpha}\left(\tilde{X} - \tilde{\mu}\right) < y\right\} = P\left\{\tilde{X} < \tilde{\mu} + \frac{y}{n^{\alpha}}\right\} \\ &= P\left\{\tilde{X} < x\right\} \text{ mit } x = \tilde{\mu} + \frac{y}{n^{\alpha}} \\ &= P\left\{N_x \geq \tfrac{n+1}{2}\right\} \text{ wobei } N_x = \#\left\{i \,\middle|\, X_i < x\right\} \sim Bi(n,p) \text{ mit } p = F(x) \\ &= P\left\{\frac{N_x - np}{\sqrt{np(1-p)}} \geq (*)\right\} \text{ wobei } (*) = \sqrt{n}\,\frac{\frac{1}{2} - p + \frac{1}{2n}}{\sqrt{p(1-p)}}. \end{aligned} \tag{15}$$

Nun ist $F(\tilde{\mu}) = \frac{1}{2}$ und

$$F(\tilde{\mu} + \varepsilon) - F(\tilde{\mu}) \;=\; \varepsilon f(\tilde{\mu}) + \tfrac{1}{2}\varepsilon^2 f'(\tilde{\mu}) + \tfrac{1}{6}\varepsilon^3 f''(\tilde{\mu}) + o\left(\varepsilon^3\right) \;=\; \tfrac{1}{6}\varepsilon^3 f''(\tilde{\mu}) + o\left(\varepsilon^3\right),$$

und somit erhalten wir

$$\sqrt{n}\left(p - \tfrac{1}{2}\right) = \sqrt{n}\left[F\left(\tilde{\mu} + \frac{y}{n^{\alpha}}\right) - F(\tilde{\mu})\right] = \sqrt{n}\left[\tfrac{1}{6}\left(\frac{y}{n^{\alpha}}\right)^3 f''(\tilde{\mu}) + o\left(n^{-3\alpha}\right)\right].$$

Falls $n^{3\alpha} = \sqrt{n}$ ist, d.h. falls $\alpha = \frac{1}{6}$ ist, so gilt

$$\sqrt{n}\left(p - \tfrac{1}{2}\right) = \tfrac{1}{6}y^3 f''(\tilde{\mu}) + o(1), \text{ d.h. } \sqrt{n}\left(p - \tfrac{1}{2}\right) \quad \xrightarrow[n\to\infty]{} \quad \tfrac{1}{6}y^3 f''(\tilde{\mu}).$$

Weiter gilt dann für $n \to \infty$

$$(*) = \sqrt{n}\,\frac{\frac{1}{2} - p + \frac{1}{2n}}{\sqrt{p(1-p)}} \quad \longrightarrow \quad -\tfrac{1}{3} y^3 f''(\tilde{\mu}),$$

und wir erhalten aus (15)

$$G_n(y) = P\left\{\frac{N_x - np}{\sqrt{np(1-p)}} \geq (*)\right\} \quad \xrightarrow[n\to\infty]{} \quad G(y) = P\left\{Z > -\tfrac{1}{3} y^3 f''(\tilde{\mu})\right\} \text{ wobei } Z \sim N(0,1).$$

Die Grenzverteilung *G* kann auch geschrieben werden als

$$G(y) = P\left\{Z < \tfrac{1}{3} y^3 f''(\tilde{\mu})\right\} = P\left\{\frac{3Z}{f''(\tilde{\mu})} < y^3\right\} = P\{Y < y\},$$

wobei $Y^3 = 3Z / f''(\tilde{\mu}) \sim N(0, \tau^2)$ mit $\tau^2 = \left(3/f''(\tilde{\mu})\right)^2$. Somit können wir schreiben

$$Y_n = n^{1/6}(\tilde{X} - \tilde{\mu}) \quad \xrightarrow[n\to\infty]{d} \quad Y, \text{ wobei } Y^3 \sim N(0,\tau^2) \text{ mit } \tau^2 = \left(\frac{3}{f''(\tilde{\mu})}\right)^2.$$

Auch diese Beziehung gilt nicht bloss für ungerade Werte von *n* sondern allgemein. Die Konvergenzgeschwindigkeit beträgt also hier nur noch $n^{1/6}$ anstelle von $\sqrt{n}$. Wir habe bisher angenommen, dass $f(\tilde{\mu}) = f'(\tilde{\mu}) = 0$ und $f''(\tilde{\mu}) > 0$. Wenn wir annehmen, dass $f(\tilde{\mu}) = f'(\tilde{\mu}) = f''(\tilde{\mu}) = f'''(\tilde{\mu}) = 0$ und $f^{(4)}(\tilde{\mu}) > 0$, so erhalten wir auf analoge Weise wie oben das Ergebnis

$$Y_n = n^{1/10}(\tilde{X} - \tilde{\mu}) \quad \xrightarrow[n\to\infty]{d} \quad Y, \text{ wobei } Y^5 \sim N(0,\tau^2) \text{ mit } \tau^2 = \left(\frac{5!}{2 f^{(4)}(\tilde{\mu})}\right)^2.$$

Die Konvergenzgeschwindigkeit beträgt jetzt nur noch $n^{1/10}$ anstelle von $\sqrt{n}$ wie im klassischen Fall.

## 6. Asymptotische Verteilung von beliebigen Quantilen

Im Abschnitt 4 haben wir die asymptotische Verteilung des Zentralwerts untersucht. Hier wollen wir die asymptotische Verteilung eines beliebigen Quantils bestimmen. Es seien $X_1, \ldots, X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion *F* und mit der Dichtefunktion *f*. Weiter sei *q* ein beliebiges Quantum mit $0 < q < 1$. Das Quantil $x_q$ ist so definiert, dass $F(x_q) = q$. Das Stichprobenquantil zum Quantum *q* definieren wir wie im Abschnitt 3; es sei $\tilde{k} = (n+1)q = k + \delta$, wobei *k* ganzzahlig und $0 \leq \delta < 1$. Dann ist $Q = X_{(k)} + \delta(X_{(k+1)} - X_{(k)})$. Wir nehmen nun wiederum vereinfachend an, dass $(n+1)q$ ganzzahlig ist, sodass $k = (n+1)q$ und $\delta = 0$; es ist dann $Q = X_{(k)}$. Nun sei $Y_n = \sqrt{n}(Q - x_q)$. Dann gilt analog zu (13)

$$\begin{aligned} G_n(y) &= P\{Y_n < y\} = P\left\{\sqrt{n}\left(Q - x_q\right) < y\right\} = P\left\{Q < x_q + \frac{y}{\sqrt{n}}\right\} \\ &= P\{Q < x\} \text{ mit } x = x_q + \frac{y}{\sqrt{n}} \\ &= P\{N_x \geq (n+1)q\} \text{ wobei } N_x = \#\{i \mid X_i < x\} \sim Bi(n,p) \text{ mit } p = F(x) \\ &= P\left\{\frac{N_x - np}{\sqrt{np(1-p)}} \geq (*)\right\} \text{ wobei } (*) = \sqrt{n}\,\frac{q - p + \frac{q}{n}}{\sqrt{p(1-p)}}. \end{aligned} \tag{16}$$

Nun ist $F(x_q) = q$ und

$$p = F(x) = F\left(x_q + \frac{y}{\sqrt{n}}\right) \quad \xrightarrow[n\to\infty]{} \quad F(x_q) = q,$$

und wir erhalten

$$\sqrt{n}\,(p-q)=\sqrt{n}\left[F\left(x_q+\frac{y}{\sqrt{n}}\right)-F\left(x_q\right)\right]=y\,\frac{F\left(x_q+\frac{y}{\sqrt{n}}\right)-F\left(x_q\right)}{y/\sqrt{n}}\ \xrightarrow[n\to\infty]{}\ y\,f\left(x_q\right).$$

Somit gilt für $n\to\infty$

$$(*)=\sqrt{n}\,\frac{q-p+\frac{q}{n}}{\sqrt{p(1-p)}}\ \longrightarrow\ -\frac{y\,f\left(x_q\right)}{\sqrt{q(1-q)}},$$

und wir erhalten aus (16)

$$G_n(y)=P\left\{\frac{N_x-np}{\sqrt{np(1-p)}}\ge(*)\right\}\ \xrightarrow[n\to\infty]{}\ G(y)=P\left\{Z>-\frac{y\,f\left(x_q\right)}{\sqrt{q(1-q)}}\right\}\ \text{wobei}\ Z\sim N(0,1).$$

Die Grenzverteilung $G$ kann auch geschrieben werden als

$$G(y)\ =\ P\left\{Z<\frac{y\,f\left(x_q\right)}{\sqrt{q(1-q)}}\right\}\ =\ P\left\{\frac{\sqrt{q(1-q)}}{f\left(x_q\right)}Z<y\right\},$$

und somit gilt

(17) $Y_n=\sqrt{n}(Q-x_q)\ \xrightarrow[n\to\infty]{d}\ N(0,\tau^2)$ mit $\tau^2=\dfrac{q(1-q)}{f^2\left(x_q\right)}$.

Diese Beziehung gilt auch, falls $(n+1)q$ nicht ganzzahlig ist (vgl. Abschnitt 3). Aus (17) folgt insbesondere, dass $Q\xrightarrow{P}x_q$ für $n\to\infty$. Falls $f(x_q)=f'(x_q)=0$ und $f''(x_q)>0$, so findet man wie im Abschnitt 5 das Ergebnis

(18) $Y_n=n^{1/6}(Q-x_q)\ \xrightarrow[n\to\infty]{d}\ Y,$ wobei $Y^3\sim N(0,\tau^2)$ mit $\tau^2=q(1-q)\left(\dfrac{3!}{f''\left(x_q\right)}\right)^2$.

## 7. Relative Effizienz von Schätzfunktionen

$X_1,\ldots,X_n$ seien unabängige Zufallsgrößen je mit der gleichen Verteilungsfunktion $F_\vartheta$ mit $\vartheta\in\Theta\ (\subset\mathbb{R})$; $\vartheta$ ist also ein eindimensionaler Parameter. Nun betrachten wir zwei verschiedene Schätzfunktionen für den Parameter $\vartheta$:

$T_n^A=T_n^A(X_1,\ldots,X_n)$: 1. Schätzfunktion für $\vartheta$;

$T_n^B=T_n^B(X_1,\ldots,X_n)$: 2. Schätzfunktion für $\vartheta$.

*Beispiel:* $\vartheta=$ Median von $F_\vartheta=$ Erwartungswert von $F_\vartheta$, $T_n^A=\bar{X}=$ arithmetische Mittel, $T_n^B=\tilde{X}=$ Zentralwert von $X_1,\ldots,X_n$.

Wir suchen nun ein Maß für die relative Güte der beiden Schätzverfahren. Dazu betrachten wir zwei Stichproben mit unterschiedlichen Umfängen $n_A$ und $n_B$:

1. Stichprobe: $X_1,\ldots,X_{n_A}\longrightarrow T_{n_A}^A\longrightarrow F_{n_A}^A=$ Verteilungsfunktion zu $T_{n_A}^A$;

2. Stichprobe: $X_1,\ldots,X_{n_B}\longrightarrow T_{n_B}^B\longrightarrow F_{n_B}^B=$ Verteilungsfunktion zu $T_{n_B}^B$.

Falls die Stichprobenumfänge $n_A$ und $n_B$ so gewählt werden können, dass die beiden Verteilungsfunktionen $F_{n_A}^A$ und $F_{n_B}^B$ identisch sind, d.h. dass $F_{n_A}^A\equiv F_{n_B}^B$, so heißt

$\dfrac{n_B}{n_A}=re(A:B)=$ relative Effizienz von Verfahren A verglichen mit Verfahren B.

Wenn beispielsweise für $n_A = 100$ und $n_B = 200$ die beiden Verteilungsfunktionen $F^A_{n_A}$ und $F^B_{n_B}$ identisch sind, so beträgt die relative Effizienz von Verfahren A verglichen mit Verfahren B

$$re(A:B) = \frac{n_B}{n_A} = \frac{200}{100} = 2.$$

Wir sagen dann, A ist doppelt so effizient wie B, oder A ist um 100% effizienter als B.

*Bemerkung:*

Leider ist es bei endlichen Stichprobenumfängen meistens nicht möglich, $n_A$ und $n_B$ so zu wählen, dass die Verteilungsfunktionen $F^A_{n_A}$ und $F^B_{n_B}$ identisch werden. Man könnte sich nun damit begnügen, die Varianz als Vergleichskriterium zu nehmen, und die Stichprobenumfänge $n_A$ und $n_B$ so wählen, dass die Varianzen $\sigma_A^2$ und $\sigma_B^2$ der beiden Schätzfunktionen $T^A_{n_A}$ und $T^B_{n_B}$ übereinstimmen; die relative Effizienz würde dann definiert wie oben als $re(A:B) = n_B/n_A$ . Tabelle 1 zeigt, dass es im allgemeinen nicht möglich ist, bei gegebenem $n_A$ einen Stichprobenumfang $n_B$ so zu finden, dass die Varianzen $\sigma_A^2$ und $\sigma_B^2$ exakt übereinstimmen; die relative Effizienz konvergiert jedoch gegen einen Grenzwert bei wachsenden Stichprobenumfängen.

*Beispiel:* Falls $X_i \sim N(\mu,1)$, so gilt:

für Verfahren $T^A_{n_A} = \bar{X}$ : $\sigma_A^2 = Var\left(\bar{X}\right) = \frac{1}{n_A}$

für Verfahren $T^B_{n_B} = \tilde{X}$ : $\sigma_B^2 = Var\left(\tilde{X}\right) \approx \frac{\pi}{2n_B}$ (falls $n_B$ groß).

Falls $n_B/n_A \approx \pi/2$, so gilt $\sigma_A^2 \approx \sigma_B^2$, und die relative Effizienz des arithmetischen Mittels verglichen mit dem Zentralwert beträgt somit bei der Normalverteilung für große Stichprobenumfänge etwa $re(\bar{X}:\tilde{X}) \approx \pi/2 = 1.571$. Im folgenden Abschnitt wollen wir den Begriff der asymptotischen relativen Effizienz genauer definieren.

## 8. Asymptotische relative Effizienz von Schätzfunktionen

Wir wollen den Begriff der asymptotischen relativen Effizienz an unserem Beispiel mit dem arihmetischen Mittel und dem Zentralwert erklären. Es seien $X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion *F* und mit der Dichtefunktion *f*. Wir nehmen an, dass $\mu = E(X_i) = med(X_i)$, dass also der Median und der Erwartungswert des Verteilungsfunktion *F* zusammenfallen; weiter sei $\sigma^2 = Var(X_i)$ mit $0 < \sigma^2 < \infty$ und für die Dichte beim Median gelte $0 < f(\mu) < \infty$. Nun betrachten wir als Schätzfunktionen für den Parameter $\mu$ die folgenden beiden Verfahren:

Verfahren A: $\bar{X} =$ arithmetisches Mittel von $X_1,\ldots,X_n$;

Verfahren B: $\tilde{X} =$ Zentralwert von $X_1,\ldots,X_n$.

Wir wissen bereits, dass die folgenden asymptotischen Beziehungen gelten:

$$\begin{aligned} &\sqrt{n}\left(\bar{X}-\mu\right) \xrightarrow[n\to\infty]{d} N\left(0,\sigma_A^2\right), \text{ wobei } \sigma_A^2 = \sigma^2 \\ &\sqrt{n}\left(\tilde{X}-\mu\right) \xrightarrow[n\to\infty]{d} N\left(0,\sigma_B^2\right), \text{ wobei } \sigma_B^2 = \frac{1}{4f^2(\mu)}. \end{aligned} \tag{19}$$

Nun wollen wir aus den beiden asymptotischen Beziehungen (19) die asymptotische relative Effizienz herleiten. Zunächst wollen wir das Ergebnis, das wir finden werden, plausibel machen. Für große Werte des Stichprobenumfangs $n$ gilt:

$$(20)\quad \begin{aligned} \bar{X} - \mu &\overset{d}{\approx} N\left(0,\ \sigma_A^2/n\right) \\ \tilde{X} - \mu &\overset{d}{\approx} N\left(0, \sigma_B^2/n\right). \end{aligned}$$

Wenn nun die Stichprobenumfänge $n_A$ und $n_B$ bei den beiden Verfahren A und B so gewählt werden, dass

$$\frac{\sigma_A^2}{n_A} = \frac{\sigma_B^2}{n_B},\ \text{d.h.}\ \frac{n_B}{n_A} = \frac{\sigma_B^2}{\sigma_B^2},$$

dann sind die beiden Grenzverteilungen nahezu identisch; daher beträgt die asymptotische relative Effizienz des arithmetischen Mittels im Vergleich zum Zentralwert

$$(21)\quad are\left(\bar{X} : \tilde{X}\right) = \lim \frac{n_B}{n_A} = \frac{\sigma_B^2}{\sigma_A^2} = \frac{1}{4\sigma^2 f^2(\mu)}.$$

Diese Herleitung kann jedoch nicht voll überzeugen; die Grenzverteilungen in (20) sind ausgeartete Verteilungen, da ja die Varianzen $\sigma_A^2/n$ und $\sigma_B^2/n$ gegen Null konvergieren, und die Limesbeziehung in (21) ist nicht klar definiert. Wir wollen jetzt die Herleitung der asymptotischen relativen Effizienz etwas genauer durchführen. Wir schreiben

Verfahren A: $n_A =$ Stichprobenumfang , $\quad \bar{X}_A =$ arithm. Mittel von $X_1,\ldots,X_{n_A}$ ;

Verfahren B: $n_B =$ Stichprobenumfang , $\quad \tilde{X}_B =$ Zentralwert von $X_1,\ldots,X_{n_B}$ .

Gemäß (19) gilt

$$(22)\quad \begin{aligned} \sqrt{n_A}\left(\bar{X}_A - \mu\right) &\xrightarrow[n\to\infty]{d} N\left(0,\sigma_A^2\right),\ \text{wobei}\ \sigma_A^2 = \sigma^2 \\ \sqrt{n_B}\left(\tilde{X}_B - \mu\right) &\xrightarrow[n\to\infty]{d} N\left(0,\sigma_B^2\right),\ \text{wobei}\ \sigma_B^2 = \frac{1}{4f^2(\mu)}. \end{aligned}$$

Anstelle von $n_A, n_B$ führen wir die beiden Variablen $n = n_A$ und $\lambda_n = n_B/n_A$ ein. Nun machen wir den Grenzübergang

$$n \to \infty \text{ und } \lambda_n\left(=\frac{n_B}{n_A}\right) \longrightarrow \lambda\left(=\lim_{n\to\infty}\frac{n_B}{n_A}\right),$$

und wir wählen den Grenzwert $\lambda$ so, dass die beiden asymptotischen Verteilungen identisch sind. Mit den neuen Variablen $n$ und $\lambda_n$ lauten die beiden Beziehungen (22)

$$(23)\quad \begin{aligned} \sqrt{n}\left(\bar{X}_A - \mu\right) &\xrightarrow[n\to\infty]{d} N\left(0,\sigma_A^2\right),\ \text{wobei}\ \sigma_A^2 = \sigma^2 \\ \sqrt{n\lambda_n}\left(\tilde{X}_B - \mu\right) &\xrightarrow[n\to\infty]{d} N\left(0,\sigma_B^2\right),\ \text{wobei}\ \sigma_B^2 = \frac{1}{4f^2(\mu)}. \end{aligned}$$

Der Vergleich der asymptotischen Verteilungen ist nur sinnvoll, wenn wir die Schätzfunktionen $\bar{X}_A$ und $\tilde{X}_B$ mit dem gleichen Faktor skalieren; daher multiplizieren wir die zweite Beziehung in (23) mit $1/\sqrt{\lambda_n}$ , und da $\lambda_n \to \lambda$ konvergiert, finden wir

$$\begin{aligned} \sqrt{n}\left(\bar{X}_A - \mu\right) &\xrightarrow[n\to\infty]{d} N\left(0,\sigma_A^2\right),\ \text{wobei}\ \sigma_A^2 = \sigma^2 \\ \sqrt{n}\left(\tilde{X}_B - \mu\right) &\xrightarrow[n\to\infty]{d} N\left(0,\sigma_B^2/\lambda\right),\ \text{wobei}\ \sigma_B^2 = \frac{1}{4f^2(\mu)}. \end{aligned}$$

Nun wählen wir den Grenzwert $\lambda$ so, dass die beiden asymptotischen Verteilungen übereinstimmen; es muss dann gelten:

$$\sigma_A^2 = \frac{\sigma_B^2}{\lambda}, \text{ d.h. } \lambda = \frac{\sigma_B^2}{\sigma_A^2}.$$

Somit haben wir das Resultat

$$are\left(\bar{X} : \tilde{X}\right) = \lim_{n\to\infty} \lambda_n = \lim \frac{n_B}{n_A} = \lambda = \frac{\sigma_B^2}{\sigma_A^2} = \frac{1}{4\sigma^2 f^2(\mu)}.$$

Damit hat sich das Ergebnis bestätigt, das wir oben auf heuristische Weise gefunden haben.

*Bemerkungen:*

a) Die asymptotische relative Effizienz zwischen dem arithmetischen Mittel und dem Zentralwert ist unabhängig von der Skalierung der Variablen $X_1,\ldots,X_n$:

$X_i$ je mit Vert.fkt. $F$ und mit Dichte $f$;

$Y_i = c\,X_i$ je mit Vert.fkt. $G$ und mit Dichte $g$.

Dann gilt

$$G(y) = F(y/c), \quad g(y) = \frac{1}{c} f(y/c), \; Var(Y_i) = \tau^2 = c^2 Var(X_i) = c^2\sigma^2, \; \nu = med(G) = c\mu,$$

und wir finden für die Y-Variablen

$$are\left(\bar{Y} : \tilde{Y}\right) = \frac{1}{4\tau^2 g^2(\nu)} = \frac{1}{4c^2\sigma^2 \, f^2(\mu)/c^2} = \frac{1}{4\sigma^2 f^2(\mu)} = are\left(\bar{X} : \tilde{X}\right).$$

b) Das Vorgehen, das wir hier am Beispiel des arithmetischen Mittels und des Zentralwerts demonstriert haben, ist typisch für den Vergleich von vielen anderen Schätzfunktionen. Es seien $T_A$ und $T_B$ zwei Schätzfunktionen für einen Parameter $\vartheta$. Dann gilt in vielen Fällen:

- Die Schätzfunktionen sind konsistent: $T_A \xrightarrow[n\to\infty]{P} \vartheta$ und $T_B \xrightarrow[n\to\infty]{P} \vartheta$.
- Die asymptotische Verteilung der Schätzfunktionen ist eine Normalverteilung:

  $$\sqrt{n}\left(T_A - \vartheta\right) \xrightarrow[n\to\infty]{d} N\left(0, \sigma_A^2\right)$$

  $$\sqrt{n}\left(T_B - \vartheta\right) \xrightarrow[n\to\infty]{d} N\left(0, \sigma_B^2\right).$$

  Die Konvergenzgeschwindigkeit beträgt bei beiden Schätzfunktionen $\sqrt{n}$.

Die asymptotische relative Effizienz ist nun nach unseren obigen Überlegungen gegeben durch

$$are\left(T_A : T_B\right) = \lim \frac{n_B}{n_A} = \frac{\sigma_B^2}{\sigma_A^2}.$$

## 9. Beispiele

In diesem Abschnitt wollen wir die asymptotische relative Effizienz zwischen arithmetischem Mittel und Zentralwert für einige konkrete Verteilungen berechnen. Es seien $X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion $F_\mu$, wobei $F_\mu(x)=F(x-\mu)$, und $f$ sei die Dichte von $F$. Die Varianz der Zufallsgrößen $X_i$ betrage $\sigma^2$. Die Dichte $f$ sei symmetrisch bezüglich des Nullpunktes; dann sind Erwartungswert und Median von $F$ gleich Null, und der Parameter $\mu$ bezeichnet den gemeinsamen Wert des Erwartungswerts und Medians von $F_\mu$. Die asymptotische relative Effizienz zwischen arithmetischem Mittel $\bar{X}$ und Zentralwert $\tilde{X}$ ist nun gegeben durch

$$are\left(\bar{X}:\tilde{X}\right)=\frac{1}{4\sigma^2 f^2(0)}.$$

**Tabelle 2:** Asymptotische relative Effizienz zwischen arithmetischem Mittel und Zentralwert

| *Verteilung* | Dichte | $f(0)$ | $\sigma^2$ | $\text{are}\left(\bar{X},\tilde{X}\right)$ |
|---|---|---|---|---|
| $\text{N}(0,1)$ | $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ | $\frac{1}{\sqrt{2\pi}}$ | 1 | $\frac{\pi}{2}\approx 1.57$ |
| Laplace | $f(x)=\frac{1}{2}e^{-\lvert x\rvert}$ | $\frac{1}{2}$ | 2 | $\frac{1}{2}$ |
| Logistisch | $f(x)=\frac{e^{-x}}{\left(1+e^{-x}\right)^2}$ | $\frac{1}{4}$ | $\frac{\pi^2}{3}$ | $\frac{12}{\pi^2}\approx 1.22$ |
| Cauchy | $f(x)=\frac{1}{\pi\left(1+x^2\right)}$ | $\frac{1}{\pi}$ | $+\infty$ | 0 |
| Rechtecksvert. | $f(x)=1$ für $0\le x\le 1$ | 1 | $\frac{1}{12}$ | 3 |
| Dreiecksvert. | $f(x)=\begin{cases}x & \text{für } 0\le x<1\\ 2-x & \text{für } 1\le x\le 2\end{cases}$ | 1 | $\frac{1}{6}$ | $\frac{3}{2}$ |
| Beta mit $a=b=0.5$ | $f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$ | $\frac{2}{\pi}$ | $\frac{1}{8}$ | $\frac{\pi^2}{2}\approx 4.93$ |
| Beta mit $a=b=1$ | für $0\le x\le 1$ | 1 | $\frac{1}{12}$ | 3 |
| Beta mit $a=b=5$ | $\sigma^2=\frac{ab}{(a+b)^2(a+b+1)}$ | $\frac{315}{128}$ | $\frac{1}{44}$ | $\approx 1.82$ |

*Bemerkung:*
Es gilt: $0\le \text{are}\left(X:X\right)\le\infty$. Die asymptotische relative Effizienz $are(\bar{X}:\tilde{X})$ kann also jeden Wert zwischen 0 und $\infty$ annehmen. Für die Cauchyverteilung ist $f(0)>0$ und die Varianz ist $\infty$, und daher ist dann $are(\bar{X}:\tilde{X})=0$; für eine Verteilung mit $f(0)=0$ und $\sigma^2<\infty$ ist $are(\bar{X}:\tilde{X})=\infty$. Bei der Normalverteilung ist $are(\bar{X}:\tilde{X})=\pi/2\approx 1.57$. Tabelle 1 zeigt, dass bei der Normalverteilung das arithmetische Mittel auch für kleine Stichproben stets genauer ist als der Zentralwert, und dass schon für $n=20$ (etwa) die asymptotischen Ergebnisse angewandt werden können; das arithmetische Mittel mit $n=100$ Beobachtungen besitzt also etwa die gleiche Genauigkeit wie der Zentralwert mit $n=157$. Bei anderen Verteilungen, z.B. bei der Laplace- oder Cauchyverteilung, ist der Zentralwert stets genauer als das arithmetische Mittel (vgl. Tabelle 1).

# 2. Asymptotische relative Effizienz zwischen zwei Testverfahren

## 1. Problemstellung

Im ersten Kapitel haben wir die asymptotische relative Effizienz von zwei Schätzfunktionen bestimmt, welche den gleichen Parameter schätzen. Hier wollen wir die asymptotische relative Effizienz von zwei Testverfahren berechnen, welche das gleiche Testproblem lösen. Wir betrachten wiederum ein einfaches Beispiel:

$X_1, \cdots, X_n$ unabhängige Zufallsgrößen je mit Verteilungsfunktion $F_\mu$ und mit Dichte $f_\mu$;

$f_\mu(x) = f(x-\mu)$, d.h. die Zufallsgröße $X_i - \mu$ hat die Dichte $f$ (Lagemodell);

$\left.\begin{aligned} \mu &= E(X_i) \\ F_\mu(\mu) &= \tfrac{1}{2} \end{aligned}\right\}$ d.h. $\mu$ = Erwartungswert von $F_\mu$ = Median von $F_\mu$;

$H_0: \mu \le \mu_0$ gegen $H_1: \mu > \mu_0$.

A) *Klassischer t-Test (bzw. z-Test):*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2;$$

$$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n};$$

verwirf $H_0$, falls $T > t_{1-\alpha}(n-1)$ (bzw. falls $T > z_{1-\alpha}$).

Dieser Test ist streng genommen nur korrekt, falls die $X_i$ normalverteilt sind. Da wir uns hier für das asymptotische Verhalten für $n \to \infty$ interessieren, so können augrund des zentralen Grenzwertsatzes auch andere Verteilungen in Betracht gezogen werden; für $n \to \infty$ gilt $t_{1-\alpha}(n-1) \to z_{1-\alpha}$.

B) *Vorzeichentest:*

$N_+ = \#\{i \mid X_i > \mu_0\}$;

$N_- = \#\{i \mid X_i < \mu_0\}$.

Da $F$ stetig ist, so gilt $P\{X_i = 0\} = 0$, und somit können Beobachtungen mit $X_i = 0$ unberücksichtigt bleiben.

Bestimme $k_r = k_r(n,\alpha) =$ rechte kritische $\alpha$-Schranke zu $Bi\left(n, \tfrac{1}{2}\right)$:

$k_r =$ kleinste ganze Zahl mit $P\left\{Bi(n,\tfrac{1}{2}) > k_r\right\} \le \alpha$.

Verwirf $H_0$, falls $N_+ > k_r$.

Nun stellt sich die Frage, welcher der beiden Tests der bessere ist.

## 2. Asymptotische Gütefunktion des t-Tests

Gütefunktion des t-Tests:

$$\beta_n(\mu) = P_\mu\left\{\frac{\bar{X}-\mu_0}{S}\sqrt{n} > t_{1-\alpha}(n-1)\right\}$$
$$= P_\mu\left\{\frac{\bar{X}-\mu}{S}\sqrt{n}+\frac{\mu-\mu_0}{S}\sqrt{n} > t_{1-\alpha}(n-1)\right\}.$$

Für $n\to\infty$ gilt $S \xrightarrow{P} \sigma$, $\frac{\bar{X}-\mu}{S}\sqrt{n} \xrightarrow{d} N(0,1)$ und $t_{1-\alpha}(n-1) \to z_{1-\alpha}$, und damit folgt

$$\beta_n(\mu) \xrightarrow[n\to\infty]{} \begin{cases} 1 & \text{für } \mu>\mu_0 \\ \alpha & \text{für } \mu=\mu_0 \\ 0 & \text{für } \mu<\mu_0. \end{cases}$$

Der t-Test ist also ein konsistenter Test. Da auch der Zeichentest konsistent ist, und wir somit beim Zeichentest die gleiche ausgeartete Grenzverteilung erhalten wie beim t-Test, so ist mit dieser Grenzfunktion keinerlei Vergleichsmöglichkeit gegeben. Wir wollen daher anstelle von $\mu$ einen neuen Maßstab (neuen Parameter) einführen, der mit wachsendem Stichprobenumfang die Auflösung verfeinert, sodass wir für $n\to\infty$ eine nicht-ausgeartete Grenzverteilung finden:

Neuer Maßstab: $\delta = \sqrt{n}(\mu-\mu_0)$;

Gütefunktion im $\delta$-Maßstab:

$$\beta_n(\delta) = P_\delta\left\{\frac{\bar{X}-\mu}{S}\sqrt{n}+\frac{\delta}{S} > t_{1-\alpha}(n-1)\right\}.$$

Für $n\to\infty$ ($\delta$ und $\alpha$ fest) gilt nun

$$\beta_n(\delta) = P_\delta\Big\{\underbrace{\frac{\bar{X}-\mu}{S}\sqrt{n}}_{\xrightarrow{d} N(0,1)} + \underbrace{\frac{\delta}{S}}_{\xrightarrow{P}\delta/\sigma} > \underbrace{t_{1-\alpha}(n-1)}_{\to z_{1-\alpha}}\Big\} \tag{1}$$

$$\longrightarrow P\left\{N(0,1) > z_{1-\alpha}-\frac{\delta}{\sigma}\right\} = P\left\{N(0,1) < -z_{1-\alpha}+\frac{\delta}{\sigma}\right\}.$$

Damit haben wir die gesuchte Grenzfunktion gefunden. Beim Vorzeichentest werden wir nun bei der Gütefunktion die gleiche Maßstabsänderung durchführen, und wir werden eine andere Grenzverteilung finden. Der Vergleich der beiden Grenzverteilungen liefert dann die gesuchte Maßzahl für die asymptotische relative Effizienz.

## 3. Asymptotische Gütefunktion des Vorzeichentests

Gütefunktion des Vorzeichentests (mit ursprünglichem Parameter $\mu$):

$\beta_n(\mu) = P_\mu\{N_+ > k_r(n,\alpha)\}$, wobei $N_+ \sim Bi(n,\vartheta)$, $\vartheta = P\{X_i > \mu_0\}$;

neuer Maßstab wie beim Vorzeichentest: $\delta = \sqrt{n}(\mu-\mu_0)$;

dann gilt

$$\vartheta = P\{X_i>\mu_0\} = 1-F_\mu(\mu_0) = 1-F(\mu_0-\mu) = 1-F\left(-\delta/\sqrt{n}\right)$$
$$= 1-F(0)+\frac{\delta}{\sqrt{n}}f(0)+O(1/n) = \tfrac{1}{2}+\frac{\delta}{\sqrt{n}}f(0)+O(1/n);$$
$$\sqrt{n}\left(\vartheta-\tfrac{1}{2}\right) \xrightarrow[\substack{n\to\infty\\ \delta \text{ fest}}]{} \delta f(0).$$

Die Gütefunktion kann geschrieben werden als

$$\beta_n(\delta) = \mathrm{P}_\delta\{N_+ > k_r(n,\alpha)\} = \mathrm{P}_\delta\left\{\frac{N_+ - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} > \frac{k_r(n,\alpha)-n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}\right\}.$$

Nun gilt für $n\to\infty$ $(\delta$ und $\alpha$ fest$)$:

$$\frac{N_+ - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \xrightarrow{d} N(0,1)$$

$$\frac{k_r(n,\alpha)-n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} = \left(\frac{k_r(n,\alpha)-n/2}{\sqrt{n}/2} - \frac{n\vartheta-n/2}{\sqrt{n}/2}\right)\frac{\sqrt{n}/2}{\sqrt{n\vartheta(1-\vartheta)}}$$

$$= \Bigg(\underbrace{\frac{k_r(n,\alpha)-n/2}{\sqrt{n}/2}}_{\to z_{1-\alpha}} - \underbrace{2\sqrt{n}\left(\vartheta-\tfrac{1}{2}\right)}_{\to 2\delta f(0)}\Bigg)\underbrace{\frac{1}{2\sqrt{\vartheta(1-\vartheta)}}}_{\to 1} \longrightarrow z_{1-\alpha} - 2\delta f(0),$$

und somit haben wir das Ergebnis

$$\beta_n(\delta) \longrightarrow P\{N(0,1) > z_{1-\alpha} - 2\delta f(0)\} = P\{N(0,1) < -z_{1-\alpha} + 2\delta f(0)\}.$$

## 4. Asymptotische relative Effizienz

Nun wollen wir die beiden asymptotischen Gütefunktionen miteinander vergleichen. Für die Gütefunktion des t-Tests (Test A) und des Zeichentests (Test B) gilt:

$$\text{(2)}\quad \begin{aligned} &\beta_n^A(\delta) \longrightarrow \beta^A(\delta) = P\{N(0,1) < -z_{1-\alpha} + c_A\delta\} \quad \text{mit } c_A = 1/\sigma, \\ &\beta_n^B(\delta) \longrightarrow \beta^B(\delta) = P\{N(0,1) < -z_{1-\alpha} + c_B\delta\} \quad \text{mit } c_B = 2f(0). \end{aligned}$$

Offensichtlich ist der t-Test (asymptotisch) besser als der Zeichentest, falls $c_A > c_B$, da dann $\beta^A(\delta) > \beta^B(\delta)$ für $\delta > 0$, d.h. für $\delta > 0$ ist dann die Ablehnwahrscheinlichkeit von $\mathrm{H}_0$ (Macht, power) beim t-Test stets größer als beim Zeichentest. Nun wollen wir eine Maßzahl für die relative Güte der beiden Tests bestimmen, die asymptotische relative Effizienz. Es sei

$n_A =$ Stichprobenumfang beim t-Test (Verfahren A);

$n_B =$ Stichprobenumfang beim Vorzeichentest (Verfahren B).

Zunächst wollen wir das Ergebnis, das wir finden werden, plausibel machen. Wir wollen die Stichprobenumfänge $n_A$ und $n_B$ so wählen, dass beide Verfahren praktisch die gleiche Gütefunktion besitzen. Das Verhältnis der beiden Stichprobenumfänge $n_B/n_A$ ist dann (im Grenzfall) die asymptotische relative Effizienz von Test A im Vergleich zu Test B. Nun gilt für große Werte von $n_A$ und $n_B$:

$$\beta_{n_A}^A(\delta_A) \approx \beta_A(\delta_A) = P\{N(0,1) < -z_{1-\alpha} + c_A\delta_A\}, \text{ wobei } \delta_A = \sqrt{n_A}(\mu-\mu_0),$$

$$\beta_{n_B}^B(\delta_B) \approx \beta_B(\delta_B) = P\{N(0,1) < -z_{1-\alpha} + c_B\delta_B\}, \text{ wobei } \delta_B = \sqrt{n_B}(\mu-\mu_0).$$

Wenn wir die beiden Gütefunktionen vergleichen wollen, so haben wir dafür zu sorgen, dass der Vergleich bei gleichem Parameter $\mu$ erfolgt. Die beiden Grenzfunktionen sind gleich, falls

$$c_A\delta_A = c_B\delta_B, \text{ d.h. falls } c_A\sqrt{n_A}(\mu-\mu_0) = c_B\sqrt{n_B}(\mu-\mu_0),$$

und daraus folgt $n_B/n_A = c_A^2/c_B^2$. Somit ist plausibel, dass die asymptotische relative Effizienz des t-Tests (Test A) im Vergleich zum Vorzeichentest (Test B) gegeben ist durch

$$are(A:B) = \lim\frac{n_B}{n_A} = \frac{c_A^2}{c_B^2} = \frac{1}{4\sigma^2 f^2(0)}.$$

Nun wollen wir die Herleitung der asymptotischen relativen Effizienz etwas genauer durchführen. Aufgrund von (2) haben wir

$$(3)\quad \begin{aligned} \beta^A_{n_A}(\delta_A) &\xrightarrow[n_A\to\infty]{} \beta^A(\delta_A) = P\{N(0,1) < -z_{1-\alpha} + c_A\delta_A\}; \\ \beta^B_{n_B}(\delta_B) &\xrightarrow[n_B\to\infty]{} \beta^B(\delta_B) = P\{N(0,1) < -z_{1-\alpha} + c_B\delta_B\}. \end{aligned}$$

Die Parameter $\delta_A$ und $\delta_B$ auf der linken Seite von (3) sind definiert als $\delta_A = \sqrt{n_A}(\mu - \mu_0)$ bzw. $\delta_B = \sqrt{n_B}(\mu - \mu_0)$. Anstelle von $n_A, n_B$ führen wir nun die Variablen $n = n_A$ und $\lambda_n = n_B/n_A$ ein, und wir machen den Grenzübergang

$$(4)\quad n \to \infty \text{ und } \lambda_n \left(= \frac{n_B}{n_A}\right) \longrightarrow \lambda \left(= \lim_{n\to\infty} \frac{n_B}{n_A}\right).$$

Dabei wollen wir den Grenzwert $\lambda$ so wählen, dass die beiden asymptotischen Gütefunktionen identisch sind (bei entsprechenden $\delta$-Werten). Da $\delta_A$ und $\delta_B$ sich auf den gleichen Parameter $\mu$ beziehen müssen, so gilt

$$(5)\quad \delta_B = \sqrt{n_B}(\mu - \mu_0) = \sqrt{n_B/n_A}\sqrt{n_A}(\mu - \mu_0) = \delta_A\sqrt{\lambda_n} \xrightarrow[n\to\infty]{} \delta_A\sqrt{\lambda}\,.$$

Beim Grenzübergang (4) ergibt sich nun aus den Beziehungen (3)

$$(6)\quad \begin{aligned} \beta^A_{n_A}(\delta_A) &\longrightarrow \beta^A(\delta_A) = P\{N(0,1) < -z_{1-\alpha} + c_A\delta_A\}; \\ \beta^B_{n_B}\left(\delta_A\sqrt{\lambda_n}\right) &\longrightarrow \beta^B\left(\delta_A\sqrt{\lambda}\right) = P\left\{N(0,1) < -z_{1-\alpha} + c_B\delta_A\sqrt{\lambda}\right\}. \end{aligned}$$

Nun wollen wir den Grenzwert $\lambda$ so wählen, daß die beiden asymptotischen Gütefunktionen übereinstimmen (bei entsprechenden $\delta$-Werten). Es gilt dann

$$c_A = c_B\sqrt{\lambda}, \quad \text{d.h.} \quad \lambda = \frac{c_A^2}{c_B^2}.$$

Für die *asymptotische relative Effizienz* (are = asymptotic relative efficiency) zwischen dem t-Test (Test A) und dem Vorzeichentest (Test B) gilt also

$$(7)\quad are(t:V) = are(A:B) = \lim_{n\to\infty} \frac{n_B}{n_A} = \lim \lambda_n = \lambda = \frac{c_A^2}{c_B^2} = \frac{1}{4\sigma^2 f^2(0)}.$$

*Bemerkungen:*

a) Wir erhalten hier auf ganz anderem Wege das gleiche Ergebnis wie bei der Berechnung der asymptotischen relativen Effizienz $are(\bar{X}:\tilde{X})$ beim Vergleich der beiden Schätzfunktionen

$\bar{X}$ = arithmetisches Mittel von $X_1, \cdots, X_n$

$\tilde{X}$ = Zentralwert von $X_1, \cdots, X_n$.

Die Tabelle 2 von Kapitel 1 mit Werten der asymptotischen relativen Effizienz bei verschiedenen Verteilungen gilt also auch für den Vergleich zwischen t-Test und Vorzeichentest. Im Fall der Normalverteilung ist $f(0) = 1/(\sigma\sqrt{2\pi})$, und die asymptotische relative Effizienz beträgt dann $are(t:V) = \pi/2 = 1.57$. Allgemein gilt $0 \le are(t:V) \le \infty$; es gibt Verteilungen mit $are(t:V) = 0$ (z.B. die Cauchyverteilung), und auch Verteilungen mit $are(t:V) = \infty$ (z.B. Verteilungen mit $f(0) = 0$ und $\sigma^2 < \infty$). Es gibt also Verteilungen, bei denen der simple Zeichentest dem klassischen t-Test überlegen ist (auch bei endlichen Stichproben).

b) Das Vorgehen, das wir hier am Beispiel des t-Tests und des Vorzeichentests demonstriert haben, ist typisch für den Vergleich von vielen anderen Testverfahren. Es seien $T_A$ und $T_B$ zwei Testverfahren zum gleichen Testproblem für einen Parameter $\vartheta$, z.B. $H_0: \vartheta \le \vartheta_0$ gegen $H_1: \vartheta > \vartheta_0$, und es seien $\beta_n^A(\vartheta)$ und $\beta_n^B(\vartheta)$ die zugehörigen Gütefunktionen. Dann gilt in vielen Fällen:

- Beide Testverfahren sind konsistent, d.h. für beide Gütefunktionen gilt
$$\beta_n(\vartheta) \xrightarrow[n\to\infty]{} \begin{cases} 1 & \text{für } \vartheta > \vartheta_0 \\ 0 & \text{für } \vartheta < \vartheta_0 \end{cases}$$
- Die asymptotische Gütefunktion ist in beiden Fällen eine Normalverteilung:
$$\beta_n^A(\delta) \xrightarrow[n\to\infty]{} \beta^A(\delta) = P\{N(0,1) > -z_{1-\alpha} + c_A\,\delta\}$$
$$\beta_n^B(\delta) \xrightarrow[n\to\infty]{} \beta^B(\delta) = P\{N(0,1) > -z_{1-\alpha} + c_B\,\delta\},$$
wobei $\delta = \sqrt{n}(\vartheta - \vartheta_0)$; die Konvergenzgeschwindigkeit beträgt also bei beiden Gütefunktionen $\sqrt{n}$.

Die asymptotische relative Effizienz zwischen den beiden Testverfahren ist dann nach unseren obigen Überlegungen gegeben durch

$$are(T_A : T_B) \;=\; \lim \frac{n_B}{n_A} \;=\; \frac{c_A^2}{c_B^2}.$$

c) Die *asymptotische relative Effizienz nach Pitman* ist (in der Notation des vorangehenden Absatzes) wie folgt definiert:

$$are(T_A : T_B) \;=\; \lim (n_B / n_A),$$

wobei

$$\vartheta_n \to \vartheta_0,\ n_A, n_B \to \infty,\ \alpha \text{ fest},\ \lim \beta_{n_A}(\vartheta_n) = \lim \beta_{n_B}(\vartheta_n) = \beta \text{ mit } 0 < \beta < 1$$

(vgl. Büning, 1994, S. 275 ff). Unter den Voraussetzungen von Absatz b) ist unsere Definition offensichtlich äquivalent zur Definition nach Pitman.

d) Die *asymptotische relative Effizienz nach Bahadur* ist wie folgt definiert:

$$are(T_A : T_B) \;=\; \lim n_B / n_A,$$

wobei

$$\alpha = \alpha_n \to 0,\ n_A, n_B \to \infty,\ \vartheta \text{ fest},\ \lim \beta_{n_A}(\vartheta) = \lim \beta_{n_B}(\vartheta) = \beta \text{ mit } 0 < \beta < 1$$

(vgl. Büning, 1994, S. 275 ff). Unter den Voraussetzungen von Absatz b) ist auch die Definition nach Bahadur äquivalent zu unserer Definition.

*Begründung:*
Gemäß Absatz b) gilt für große Stichprobenumfänge $n_A, n_B$:

$$\beta_{n_A}^A(\delta_A) \;\approx\; \beta^A(\delta_A) = P\{N(0,1) > -z_{1-\alpha} + c_A\,\delta_A\}, \quad \text{wobei } \delta_A = \sqrt{n_A}(\vartheta - \vartheta_0);$$
$$\beta_{n_B}^B(\delta_B) \;\approx\; \beta^B(\delta_B) = P\{N(0,1) > -z_{1-\alpha} + c_B\,\delta_B\}, \quad \text{wobei } \delta_B = \sqrt{n_B}(\vartheta - \vartheta_0).$$

Nun sei $\vartheta$ fest und $\alpha = \alpha_n \to 0$; dabei sollen die Stichprobenumfänge $n_A, n_B$ so gewählt werden, dass $\beta^A(\delta_A) = \beta^B(\delta_B)$ bei festem $\vartheta$. Dann muss gelten

$$-z_{1-\alpha_n} + c_A\sqrt{n_A}(\vartheta - \vartheta_0) \;=\; -z_{1-\alpha_n} + c_B\sqrt{n_B}(\vartheta - \vartheta_0), \text{ d.h. } c_A\sqrt{n_A} = c_B\sqrt{n_B},$$

und beim Grenzübergang $n, n_A, n_B \to \infty$ ergibt sich $\lim(n_B/n_A) = c_A^2/c_B^2$.

## 5. Vergleich des t-Tests mit dem Zeichentest bei endlichem Stichprobenumfang

Wir wollen prüfen, wieweit die asymptotischen Ergebnisse zum Vergleich des t-Tests mit dem Zeichentest schon für endliche Stichproben gültig sind. $X_1,\ldots,X_n$ seien unabhängige Zufallsgrößen je mit der gleichen stetigen Verteilungsfunktion *F*, und es sei $\mu = E(X_i) = med(X_i)$; $x_1,\ldots,x_n$ seien die zugehörigen Realisationen. Wir betrachten das Testproblem $\mathrm{H}_0\colon \mu \le 0$ gegen $\mathrm{H}_1\colon \mu > 0$. Wir möchten wissen, ob bei einem Stichprobenumfang von $n = 30$ der t-Test oder der Vorzeichentest die bessere Wahl ist.

*t-Test zum Signifikanzniveau* $\alpha = 0.05$

Berechne die Teststatistik

$$T = \frac{\bar{X}}{S}\sqrt{n}, \text{ wobei } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{und} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Verwirf $\mathrm{H}_0$, falls $T > t_{1-\alpha}(n-1) = t_{0.95}(29) = 1.69913$.

*Gütefunktion des t-Tests* bei der Normalverteilung $N(\mu,\sigma^2)$:

$$\beta(\mu) = P_\mu\left\{T > t_{1-\alpha}(n-1)\right\} = P_\mu\left\{\frac{\bar{X}}{S}\sqrt{n} > t_{1-\alpha}(n-1)\right\}$$

$$= P_\mu\left\{\frac{\left(\bar{X}-\mu\right)/\sigma + \mu/\sigma}{S/\sigma}\sqrt{n} > t_{1-\alpha}(n-1)\right\} = P\left\{t(n-1,\delta) > t_{1-\alpha}(n-1)\right\},$$

wobei hier $t(n-1,\delta)$ eine Zufallsgröße mit einer nicht-zentralen t-Verteilung mit den Parametern $n-1$ und $\delta = (\mu/\sigma)\sqrt{n}$ bezeichnet. Falls die Verteilungsfunktion *F* der Zufallsgrößen $X_1,\ldots,X_n$ keine Normalverteilung ist, so läßt sich die Gütefunktion des t-Tests nicht so einfach berechnen; wir können jedoch in diesen Fällen die Werte der Gütefunktion durch Simulation (Monte-Carlo-Verfahren) schätzen.

*Zeichentest zum Signifikanzniveau* $\alpha = 0.05$

Berechne die Teststatistik $N_+ = \#\left\{i \middle| X_i > 0\right\}$; bestimme die rechte kritische $\alpha$-Schranke $k_r$ zur Binomialverteilung $Bi(n,\frac{1}{2})$: $k_r =$ kleinste ganze Zahl mit $P\left\{Bi(n,\frac{1}{2}) > k_r\right\} \le \alpha$; es gilt

$$P\left\{Bi(n,\tfrac{1}{2}) \ge 19\right\} = 0.1002 > \alpha = 0.05 \text{ und } P\left\{Bi(n,\tfrac{1}{2}) > 19\right\} = 0.0494 < \alpha = 0.05,$$

und daher ist $k_r = 19$. Verwirf $\mathrm{H}_0$, falls $N_+ > k_r = 19$.

*Gütefunktion des Zeichentests:*

$$\beta(\mu) = P_\mu\left\{N_+ > k_r\right\} = P\left\{Bi(n,p_+) > k_r\right\}, \text{ wobei } p_+ = P\left\{X_i > 0\right\}.$$

Die exakte Gütefunktion des Zeichentests kann also für beliebige Verteilungen berechnet werden. Wir werden jedoch aus Neugierde auch in diesen Fällen zusätzlich die Gütefunktion durch Simulation schätzen.

Die Tabellen 1a bis 1c zeigen die Ergebnisse unseres Vergleichs bei drei verschiedenen Verteilungen:

a) Normalverteilung $N(\mu,1)$ mit Dichte $\quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$

b) Laplaceverteilung $L(\mu,1)$ mit Dichte $\quad f(x) = \frac{1}{2} e^{-|x-\mu|}$

c) Cauchyverteilung $C(\mu,1)$ mit Dichte $\quad f(x) = \frac{1}{\pi\left(1+(x-\mu)^2\right)}$.

**Tabelle 1a:** Werte der Gütefunktion bei einer Normalverteilung $N(\mu,1)$ für $n=30$ und $\alpha=0.05$; hier ist $\delta=\mu\sqrt{n}$ und $p_+=P\{X_i>0\}=P\{N(0,1)<\mu\}$.

(1) (2) (3) (4) (5) (6) (7)

| $\mu$ | $\delta$ | $p_+$ | Exakt | | Simulation | |
|---|---|---|---|---|---|---|
| | | | t-Test | Zeichen-Test | t-Test | Zeichentest |
| 0 | 0 | 0 | 0.05 | 0.0494 | 487 | 492 |
| 0.25 | 1.369 | 0.5987 | 0.3791 | 0.2865 | 3700 | 2875 |
| 0.5 | 2.739 | 0.6915 | 0.8483 | 0.6951 | 8416 | 6907 |
| 0.75 | 4.108 | 0.7734 | 0.9910 | 0.9415 | 9896 | 9403 |
| 1 | 5.477 | 0.8413 | 0.9999 | 0.9954 | 9998 | 9956 |

**Tabelle 1b:** Werte der Gütefunktion bei einer Laplaceverteilung $L(\mu,1)$ für $n=30$ und $\alpha=0.05$; hier ist $p_+=P\{X_i>0\}=P\{L(0,1)<\mu\}$.

(1) (2) (3) (4) (5) (6)

| $\mu$ | $p_+$ | Exakt | | Simulation | |
|---|---|---|---|---|---|
| | | t-Test | Zeichen-Test | t-Test | Zeichentest |
| 0 | 0.5 | ? | 0.0494 | 529 | 499 |
| 0.25 | 0.6106 | ? | 0.3336 | 2554 | 3303 |
| 0.5 | 0.6967 | ? | 0.7169 | 6066 | 7123 |
| 0.75 | 0.7638 | ? | 0.9244 | 8762 | 9224 |
| 1 | 0.8161 | ? | 0.9859 | 9755 | 9857 |

**Tabelle 1c:** Werte der Gütefunktion bei einer Cauchyverteilung $C(\mu,1)$ für $n=30$ und $\alpha=0.05$; hier ist $p_+=P\{X_i>0\}=P\{C(0,1)<\mu\}$.

(1) (2) (3) (4) (5) (6)

| $\mu$ | $p_+$ | Exakt | | Simulation | |
|---|---|---|---|---|---|
| | | t-Test | Zeichen-Test | t-Test | Zeichentest |
| 0 | 0.5 | ? | 0.0494 | 347 | 499 |
| 0.25 | 0.5780 | ? | 0.2136 | 741 | 2069 |
| 0.5 | 0.6476 | ? | 0.4967 | 1355 | 4911 |
| 0.75 | 0.7048 | ? | 0.7495 | 2132 | 7453 |
| 1 | 0.7500 | ? | 0.8943 | 2880 | 8916 |
| 1.25 | 0.7852 | ? | 0.9585 | 3589 | 9561 |
| 1.5 | 0.8128 | ? | 0.9840 | 4253 | 9827 |

Bei der Normalverteilung (vgl. Tabelle 1a) ist der t-Test dem Zeichentest deutlich überlegen; so beträgt etwa für $\mu=0.5$ die Ablehnwahrscheinlichkeit (Wahrscheinlichkeit, dass die Nullhypothese abgelehnt wird) beim t-Test 0.8483 während sie bei Zeichentest nur 0.6951 beträgt. In einer Simulation mit $M=10000$ Stichproben führte der t-Test (Zeichentest) in 8416 (6907) Fällen zur Ablehnung; wenn wir aufgrund dieser Häufigkeiten die tatsächlichen Ablehnwahrscheinlichkeiten schätzen und ein Vertrauensintervall zum Sicherheitsgrad von $0.95$ $(=1-2\gamma)$ bestimmen, so erhalten wir mit den klassichen Formeln zur Bestimmung von Vertrauensgrenzen bei einer Binomialverteilung:

$$\beta = \hat{\beta} \pm z_{1-\gamma}\sqrt{\frac{\hat{\beta}(1-\hat{\beta})}{M}}\,;$$

nun gilt $\sqrt{\hat{\beta}(1-\hat{\beta})} \le \frac{1}{2}$ und für $\gamma = 0.025$ ist $z_{1-\gamma} = z_{0.975} = 1.9600 \approx 2$, und somit haben wir

$$z_{1-\gamma}\sqrt{\frac{\hat{\beta}(1-\hat{\beta})}{M}} < \frac{1}{\sqrt{M}} = 0.01\,.$$

Wenn wir also die Häufigkeiten in Spalte 6 und 7 von Tabelle 1a durch $M = 10\,000$ dividieren, so erhalten wir Schätzwerte für die exakte Gütefunktion, welche mit einem Sicherheitsgrad von 95% um höchstens 0.01 vom wahren Wert abweichen. Der Vergleich mit den exakten Werten bestätigt die Schätzwerte.

Tabelle 1b enthält die Ergebnisse des Vergleichs bei der Laplaceverteilung (zweiseitige Exponentialverteilung). Die exakte Gütefunktion des t-Tests kann hier nicht berechnet werden; die Zufallsgrößen $\overline{X}$ und $S^2$ sind ja nicht mehr unabhängig wie im Falle der Normalverteilung. Die Simulationsergebnisse zeigen, dass jetzt der Zeichentest dem t-Test eindeutig überlegen ist.

Auch bei der Cauchyverteilung (vgl. Tabelle 1c) zeigen die Simulationsergebnisse, dass der Zeichentest dem t-Test deutlich überlegen ist. Es ist jedoch bemerkenswert, dass der t-Test für $\mu = 0$ das Signifikanzniveau $\alpha = 0.05$ nicht verletzt sondern deutlich unterschreitet (Schätzwert: 0.0347).

Wie kommen die Simulationsergebnisse in den Tabellen 1a bis 1c zustande? Wir nehmen an, dass die Zufallsgröße *X* eine Laplaceverteilung (zweiseitige Exponentialverteilung) $L(\mu,\sigma)$ besitzt mit der Dichte

$$f(x) = \tfrac{1}{2}e^{-|x-\mu|/\sigma},\ x \in \mathbb{R}\,,$$

und wir betrachten den speziellen Fall $\mu = 0.5$ und $\sigma = 1$. Wir möchten wissen, ob für $n = 30$ der t-Test oder der Zeichentest die bessere Wahl ist. Um diese Frage zu beantworten, wollen wir eine Simulationsstudie durchführen. Wir erzeugen mit Minitab $M = 10\,000$ Stichproben je mit Umfang $n = 30$ aus einer Laplaceverteilung $L(\mu,1)$ mit $\mu = 1/2$, und führen für jede der $M = 10\,000$ Stichproben sowohl den klassischen t-Test als auch den Zeichentest durch (Signifikanzniveau $\alpha = 0.05$):

1. Wähle als Startwert des Zufallszahlengenerators den Wert 123.
2. Erzeuge in den Spalten 11 bis 40 des Arbeitsblattes je $M = 10\,000$ Zufallszahlen aus $L(\mu,1)$ mit $\mu = 1/2$ (vgl. Tabelle 2).
3. *t-Test:* Berechne für jede Stichprobe das arithmetische Mittel $\overline{x}$, die empirische Standardabweichung *s* und den Wert der Teststatistik: $t = \frac{\overline{x}}{s}\sqrt{n}$ (vgl. Spalte 1 bis 3). Prüfe, ob $t > t_{1-\alpha}(n-1) = t_{0.95}(29) = 1.69913$. Die Antwort findet sich in Spalte 4.
4. *Zeichentest:* Berechne für jede Stichprobe die Anzahl $N_+$ der positiven Werte (Spalte 5); dazu bestimmen wir die 0/1-Werte in den Spalten 41 bis 70 (0 = negativ, 1 = positiv). Prüfe, ob $N_+ > k_r = 19$, wobei $k_r = 19$ die rechte kritische $\alpha$-Schranke der Binomialverteilung $Bi(n,\frac{1}{2})$ bezeichnet. Die Testentscheidung ist in Spalte 6 zu finden.

**Tabelle 2:** Simulationsstudie mit $M = 10\,000$ Stichproben aus einer Laplace-Verteilung $L\left(\frac{1}{2},1\right)$

| (0) | (1) | (2) | (3) | (4) | (5) | (6) | (11) | ⋯ | (40) | (41) | ⋯ | (70) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nr. | t-Test | | | | Zeichentest | | Beobachtungen | | | 0/1-Werte | | |
| | $\overline{x}$ | $s$ | $t$ | $H_1$? | $N_+$ | $H_1$? | $x_1$ | ⋯ | $x_{30}$ | $y_1$ | ⋯ | $y_{30}$ |
| 1 | 0.37101 | 1.37883 | 1.47378 | 0 | 19 | 0 | 0.64079 | … | -1.3080 | 1 | … | 0 |
| 2 | 0.03373 | 1.42418 | 0.12972 | 0 | 19 | 0 | 0.31508 | … | -0.2232 | 1 | … | 0 |
| 3 | 0.82887 | 1.30051 | 3.49088 | 1 | 22 | 1 | -0.88260 | … | 0.1295 | 0 | … | 1 |
| 4 | 0.65054 | 1.45029 | 2.45687 | 1 | 22 | 1 | 1.67442 | … | 0.7213 | 1 | … | 1 |
| 5 | 0.41941 | 0.98456 | 2.33322 | 1 | 19 | 0 | 1.12817 | … | -0.7201 | 1 | … | 0 |
| 6 | 0.00896 | 1.23872 | 0.03962 | 0 | 16 | 0 | 0.47935 | … | 0.2135 | 1 | … | 1 |
| 7 | 0.39281 | 1.49749 | 1.43675 | 0 | 23 | 1 | 1.53114 | … | 3.5646 | 1 | … | 1 |
| 8 | 0.47554 | 1.28752 | 2.02299 | 1 | 22 | 1 | 0.74455 | … | 0.8970 | 1 | … | 1 |
| 9 | 0.39541 | 1.13347 | 1.91075 | 1 | 21 | 1 | 0.71832 | … | 1.2125 | 1 | … | 1 |
| 10 | 0.98364 | 1.49444 | 3.60511 | 1 | 25 | 1 | 0.90279 | … | -1.0527 | 1 | … | 0 |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 9991 | 0.57491 | 1.29471 | 2.43214 | 1 | 22 | 1 | -0.08564 | … | 0.8224 | 0 | … | 1 |
| 9992 | 0.08049 | 1.52525 | 0.28904 | 0 | 16 | 0 | -0.69618 | … | 0.2219 | 0 | … | 1 |
| 9993 | 0.63200 | 1.34739 | 2.56911 | 1 | 23 | 1 | 0.77270 | … | 0.1727 | 1 | … | 1 |
| 9994 | 0.32637 | 1.70998 | 1.04541 | 0 | 20 | 1 | -0.89625 | … | 0.1267 | 0 | … | 1 |
| 9995 | 0.94774 | 1.85888 | 2.79251 | 1 | 22 | 1 | 0.74716 | … | -0.0647 | 1 | … | 0 |
| 9996 | 0.88811 | 1.39718 | 3.48158 | 1 | 24 | 1 | 1.31647 | … | 0.7731 | 1 | … | 1 |
| 9997 | 0.65453 | 1.57324 | 2.27875 | 1 | 21 | 1 | 1.64722 | … | 0.2508 | 1 | … | 1 |
| 9998 | 0.41739 | 1.69666 | 1.34743 | 0 | 19 | 0 | -0.08139 | … | 0.7768 | 0 | … | 1 |
| 9999 | 0.59409 | 0.83603 | 3.89218 | 1 | 22 | 1 | 1.62783 | … | 0.6911 | 1 | … | 1 |
| 10000 | 0.49580 | 1.64452 | 1.65132 | 0 | 22 | 1 | 0.61088 | … | 1.1070 | 1 | … | 1 |
| | | | | 6066 | | 7123 | | | | | | |

*Minitab-Makro* `MINITAB1.MAC` *zur Berechnung der Daten in Tabelle 1*

```
gmacro
note
# Einstichprobenproblem;
#   Vergleich des t-Tests mit dem Zeichentest
#   hier: n=30, Laplaceverteilung L(0.5,1)
#

name c1 'x-quer' c2 's' c3 't' c4 'H1?' c6 'N_plus' c7 'H1??'
Base 123.
Random 10000 c11-c40;
  Laplace 0.5 1.
RMean C11-C40 c1.
RStDev C11-C40 c2.
let c3 = sqrt(30)*C1/c2
Code (-100:0) 0 (0:100) 1 c11-c40 c41-c70
RSum c41-c70 c6
InvCDF 0.95 k1;
  T 29.
prin k1
Code (-100:k1) 0 (k1:100) 1 c3 c4
tally c4
InvCDF 0.95 k2;
  Binomial 30 0.5.
InvCDF 0.95;
  Binomial 30 0.5.
prin k2
Code (0:k2) 0 (k2:30) 1 c6 c7
tally c7
let c8 = c4+c7
tally c8
endmacro
```

Wir stellen fest, dass der t-Test die Nullhypothese in 6066 Fällen, der Zeichentest jedoch in 7123 Fällen ablehnen kann; diese zwei Häufigkeiten finden sich in Tabelle 1b in der Zeile $\mu = 0.5$. Der Zeichentest ist somit trennschärfer als der t-Test. Analog werden die übrigen Simulationsergebnisse in den Tabellen 1a bis 1c bestimmt.

Aufgrund der beiden Häufigkeiten 6066 und 7123 finden wir die folgenden Schätzwerte für den exakten Wert der Gütefunktion bei $\mu = \frac{1}{2}$:

t-Test: $\beta(\mu) = 0.6066 \pm 0.01$ (Sicherheitsgrad 0.95)

Zeichen-Test: $\beta(\mu) = 0.7123 \pm 0.01$ (Sicherheitsgrad 0.95)

Es scheint plausibel, dass die Ablehnwahrscheilichkeit beim Zeichentest eindeutig größer ist als beim t-Test. Es stellt sich jedoch die Frage, ob die beiden Schätzwerte für die Gütefuktion beim t-Test und beim Zeichentest tatsächlich signifikant verschieden sind. Da wir die beiden Häufigkeiten 6066 und 7123 aufgrund der gleichen $M = 10\,000$ Stichproben bestimmt haben, so müssen wir die stochastische Abhängigkeit der beiden Häufigkeiten berücksichtigen. Dazu zählen wir die Stichproben, bei denen keiner der beiden Tests, nur einer der beiden Tests oder beide Tests die Nullhypothese ablehnen. Das Ergebnis findet sich in Tabelle 3. Bei 2218 Stichproben kann keiner der beiden Test die Nullhypothese ablehnen, in 1716 Fällen führt nur der t-Test zur Ablehnung, in 659 Fällen nur der Zeichentest, und in 5407 Fällen lehnen beide Test ab. Der korrekte Test für die Frage, ob die beiden Ablehnwahrscheinlichkeiten verschieden sind ist der Test von McNemar, den wir im Kapitel 5 behandeln werden. Für die Daten in Tabelle 3 sind die beiden Ablehnwahrscheinlichkeiten hochsignifikant verschieden (p-Wert bei zweiseitiger Fragestellung $< 10^{-100}$).

**Tabelle 3:** Ergebnisse zur Simulationsstudie in Tabelle 2; t = t-Test, Z = Zeichentest

| t \ Z | $H_0$ | $H_1$ | $\Sigma$ |
|---|---|---|---|
| $H_0$ | 2218 | 1716 | 3934 |
| $H_1$ | 659 | 5407 | 6066 |
| $\Sigma$ | 2877 | 7123 | 10 000 |

# 3. Median-Verfahren

## 1. Allgemeine Definition des Medians

Es sei $X$ eine (eindimensionale) Zufallsgröße mit einer beliebigen Verteilungsfunktion $F$ (stetig, diskret, oder Mischtyp). Ein Wert $\mu \in \mathbb{R}$ heißt *Median* von $F$, falls

$$F(\mu - 0) \le \tfrac{1}{2} \quad \text{und} \quad F(\mu + 0) \ge \tfrac{1}{2},$$

d.h. falls

$$P\{X < \mu\} \le \tfrac{1}{2} \quad \text{und} \quad P\{X > \mu\} \le \tfrac{1}{2}.$$

Falls die Verteilungsfunktion $F$ stetig ist, so gilt:

$$F(\mu) = \tfrac{1}{2}, \text{ d.h. } P\{X < \mu\} = P\{X > \mu\} = \tfrac{1}{2}.$$

Man beachte, daß der Median nicht bei jeder Verteilung eindeutig festgelegt ist; falls ein Intervall $[a,b]$ existiert mit $a < b$ und $F(a) = F(b) = \frac{1}{2}$, so ist nach unserer Definition jeder Wert $\mu \in [a,b]$ ein Median.

## 2. Vertrauensgrenzen für den Median

$X_1, \ldots, X_n$ unabhängige Zufallsgrößen je mit der unbekannten Verteilungsfunktion $F$;
$\mu = \text{med}(F) =$ Median von $F$ ;
$x_1, \ldots, x_n$ Realisationen zu $X_1, \ldots, X_n$ ;
*Gesucht:* Vertrauensgrenzen für $\mu$.

*Lösung:*

1. Ordne die Beobachtungen $x_1, \ldots, x_n$ nach ihrer Größe: $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$.

2. Bestimme zu gegebenen $\alpha$ $(0 < \alpha < \frac{1}{2})$ die linke und rechte kritische Schranke zu $Bi(n, \frac{1}{2})$:

   $k_\ell = k_\ell(n,\alpha) =$ größte ganze Zahl mit $P\left\{Bi(n,\tfrac{1}{2}) < k_\ell\right\} \le \alpha$;
   $k_r = k_r(n,\alpha) =$ kleinste ganze Zahl mit $P\left\{Bi(n,\tfrac{1}{2}) > k_r\right\} \le \alpha$.

   Aus Symmetriegründen gilt dann $k_\ell + k_r = n$. Falls $P\left\{Bi(n,\tfrac{1}{2}) = 0\right\} = (\tfrac{1}{2})^n > \alpha$, so ist $k_\ell = 0$ und $k_r = n$.

   *Beispiel:* Für $\alpha = 0.01$ ist $(\tfrac{1}{2})^n > \alpha$ falls $n \le 6$, und somit ist dann $k_\ell = 0$ und $k_r = n$.

3. Dann definieren wir die Vertrauensgrenzen als

   $$\mu_\ell = \mu_\ell(x_1, \ldots, x_n, \alpha) = x_{(k_\ell)} \quad \text{und} \quad \mu_r = \mu_r(x_1, \ldots, x_n, \alpha) = x_{(k_r+1)}.$$

   Da bei den geordneten Beobachtungen $x_{(i)}$ der Index $i$ zwischen 1 und $n$ liegen muß, ist unsere Definition der Vertrauensgrenzen nur sinnvoll, falls $k_\ell > 0$ und $k_r < n$, d.h. falls $(\tfrac{1}{2})^n \le \alpha$. Wenn $n$ zu klein ist, versagt unsere Methode (z.B. für $n \le 6$ falls $\alpha = 0.01$). Man beachte auch, daß bei der Binomialverteilung $Bi(n,\tfrac{1}{2})$ für symmetrisch gelegene Wertepaare $(0,n), (1,n-1), \ldots, (k_\ell, k_r), \ldots$ die Indexsume jeweils $n$ beträgt, während sie bei den Ordnungsstatistiken $x_{(i)}$ für symmetrisch gelegene Paare $(1,n), (2,n-1), \ldots, (k_\ell, k_r+1), \ldots$ jeweils $n+1$ beträgt.

Es gilt nun die klassische Konfidenzinterpretation

1) $P_F\{\mu_\ell(X_1,\ldots,X_n,\alpha) \le \mu\} \ge 1-\alpha$

2) $P_F\{\mu \le \mu_r(X_1,\ldots,X_n,\alpha)\} \ge 1-\alpha$

3) $P_F\{\mu_\ell(X_1,\ldots,X_n,\alpha) \le \mu \le \mu_r(X_1,\ldots,X_n,\alpha)\} \ge 1-2\alpha$

für jede beliebige unbekannte Verteilungsfunktion $F$, d.h. $\mu_\ell$ und $\mu_r$ sind Vertrauensgrenzen für $\mu$ zum Sicherheitsgrad $1-\alpha$ (bzw. $1-2\alpha$). Dabei sind keinerlei Voraussetzungen für $F$ notwendig, auch keine Stetigkeit. Daher nennt man dieses Verfahren ein *verteilungsfreies* Verfahren.

*Beweis:*

Es sei

$$p_- = P\{X_i < \mu\} = F(\mu-0) \left(\le \tfrac{1}{2}\right)$$

$$p_+ = P\{X_i > \mu\} = 1-F(\mu+0) \left(\le \tfrac{1}{2}\right)$$

$$p_0 = P\{X_i = \mu\} = F(\mu+0)-F(\mu-0).$$

Weiter sei

$$N_- = \#\{i \mid X_i < \mu\} = \text{ Anzahl der Beobachtungen } X_i \text{ mit } X_i < \mu$$

$$N_+ = \#\{i \mid X_i > \mu\}$$

$$N_0 = \#\{i \mid X_i = \mu\}.$$

Es gilt $N_- \sim Bi(n, p_-)$, $N_+ \sim Bi(n, p_+)$. $N_0 \sim Bi(n, p_0)$.

$1^o$ Nun sei $\mu_\ell = \mu_\ell(X_1,\ldots,X_n,\alpha) = X_{(k_\ell)}$. Dann gilt

$$\{\mu < \mu_\ell\} = \{\mu < X_{(k_\ell)}\} = \text{"höchstens } k_\ell - 1 \text{ Beobachtungen sind } \le \mu\text{"}$$
$$= \{N_- + N_0 < k_\ell\}.$$

Nun gilt $N_- + N_0 \sim Bi(n, p_- + p_0)$ mit $p_- + p_0 \ge \frac{1}{2}$. Da die Wahrscheinlichkeit $P\{Bi(n,\vartheta) < k\}$ als Funktion von $\vartheta$ abnehmend ist in $\vartheta$ bei festem $k$ ($k = 1,\ldots,n$), so gilt

$$P\{\mu < \mu_\ell\} = P\{\mu < X_{(k_\ell)}\} = P\{N_- + N_0 < k_\ell\}$$
$$= P\{Bi(n, p_- + p_0) < k_\ell\} \le P\left\{Bi\left(n, \tfrac{1}{2}\right) < k_\ell\right\} \le \alpha.$$

Daraus ergibt sich $P\{\mu_\ell \le \mu\} \ge 1-\alpha$, und damit ist die Behauptung 1) bewiesen.

$2^o$ Nun sei $\mu_r = \mu_r(X_1,\ldots,X_n,\alpha) = X_{(k_r+1)}$. Dann gilt

$$\{\mu > \mu_r\} = \{\mu > X_{(k_r+1)}\} = \text{"mindestens } k_r + 1 \text{ Beobachtungen sind } < \mu\text{"}$$
$$= \{N_- > k_r\}.$$

Nun gilt $N_- \sim Bi(n, p_-)$ mit $p_- \le \frac{1}{2}$. Da die Wahrscheinlichkeit $P\{Bi(n,\vartheta) > k\}$ als Funktion von $\vartheta$ wachsend ist in $\vartheta$ bei festem $k$ ($k = 0,\ldots,n-1$), so gilt

$$P\{\mu > \mu_r\} = P\{\mu > X_{(k_r+1)}\} = P\{N_- > k_r\}$$
$$= P\{Bi(n, p_-) > k_r\} \le P\left\{Bi\left(n, \tfrac{1}{2}\right) > k_r\right\} \le \alpha.$$

Daraus ergibt sich $P\{\mu \le \mu_r\} \ge 1-\alpha$, und damit ist die Behauptung 2) bewiesen.

$3^o$ Da $\mu_\ell \le \mu_r$ für $\alpha \le \frac{1}{2}$, so gilt

$$\underbrace{\{\mu_\ell \le \mu \le \mu_r\}^c}_{=E} = \underbrace{\{\mu < \mu_\ell\}}_{=E_1} \cup \underbrace{\{\mu > \mu_r\}}_{=E_2},$$

und die beiden Ereignisse $E_1$ und $E_2$ sind disjunkt. Daher gilt $P(E) = P(E_1) + P(E_2) \le 2\alpha$ und somit $P\{\mu_\ell \le \mu \le \mu_r\} = P\left(E^c\right) \ge 1 - 2\alpha$. Damit ist die Behauptung 3) bewiesen.

*Ergänzung:*

Für $\alpha = \frac{1}{2}$ gilt bei ungeradem *n*

$$P\left\{Bi(n,\tfrac{1}{2}) \le \tfrac{1}{2}(n-1)\right\} = P\left\{Bi(n,\tfrac{1}{2}) > \tfrac{1}{2}(n-1)\right\} = \tfrac{1}{2},$$

und daher ist dann $k_\ell = k_r + 1 = \frac{1}{2}(n+1)$, und die beiden Vertrauensgrenzen fallen zusammen mit dem Zentralwert:

$$\mu_\ell = \mu_r = X_{(k_\ell)} = Ze = \text{Zentralwert von } X_1,\ldots,X_n.$$

Die Konfidenzeigenschaft besagt nun

$$P\{Ze \le \mu\} \ge \tfrac{1}{2} \quad \text{und} \quad P\{Ze \ge \mu\} \ge \tfrac{1}{2}.$$

Damit ist der Zentralwert bei ungeradem *n* eine mediantreue Schätzung für $\mu = med(F)$ bei beliebiger Verteilungsfunktion *F*. Für $\alpha = \frac{1}{2}$ und *n* gerade gilt

$$P\left\{Bi(n,\tfrac{1}{2}) < \tfrac{1}{2}n\right\} = P\left\{Bi(n,\tfrac{1}{2}) > \tfrac{1}{2}n\right\} < \tfrac{1}{2},$$

und daher ist dann $k_\ell = k_r = n/2$; die beiden Vertrauensgrenzen $\mu_\ell = X_{(n/2)}$ und $\mu_r = X_{(n/2+1)}$ sind somit die Endpunkte des zentralen Intervalls, und der Zentralwert *Ze* wird üblicherweise als Mittelpunkt dieses Intervalls definiert: $Ze = \frac{1}{2}(X_{(n/2)} + X_{(n/2+1)})$. Wir werden unten (im Abschnitt über randomisierte Vertrauensgrenzen) sehen, daß bei geradem *n* der Zentralwert im allgemeinen nicht keine mediantreue Schätzung für $\mu = med(F)$ liefert.

## 3. Vertrauensgrenzen für beliebige Quantile

Es sei *X* eine (eindimensionale) Zufallsgröße mit einer beliebigen Verteilungsfunktion *F* (stetig, diskret, oder Mischtyp). Ein Wert $x_q$ $(0 < q < 1)$ heißt *q*-Quantil von *F*, falls

$$F(x_q - 0) \le q \quad \text{und} \quad F(x_q + 0) \ge q,$$

d.h. falls

$$P\{X < x_q\} \le q \quad \text{und} \quad P\{X > x_q\} \le 1 - q.$$

Falls die Verteilungsfunktion *F* stetig ist, so gilt:

$$F(x_q) = q, \text{ d.h. } P\{X < x_q\} = q \text{ und } P\{X > x_q\} = 1 - q.$$

Nun seien $X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit der unbekannten Verteilungsfunktion *F*, $x_q$ ein *q*-Quantil von *F*, $x_1,\ldots,x_n$ Realisationen zu $X_1,\ldots,X_n$.

*Gesucht:* Vertrauensgrenzen für $x_q$ zum Sicherheitsgrad $1-\alpha$ (bzw. $1-2\alpha$).

*Lösung:*

1. Ordne die Beobachtungen $x_1,\ldots,x_n$ nach ihrer Größe: $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$.
2. Bestimme zu gegebenen $\alpha$ $(0 < \alpha < \frac{1}{2})$ die kritischen Schranken zu $Bi(n,q)$:

   $k_\ell = k_\ell(n,q,\alpha) =$ größte ganze Zahl mit $P\{Bi(n,q) < k_\ell\} \le \alpha$;
   $k_r = k_r(n,q,\alpha) =$ kleinste ganze Zahl mit $P\{Bi(n,q) > k_r\} \le \alpha$.

3. Dann definieren wir die Vertrauensgrenzen als

$$\mu_\ell = \mu_\ell(x_1,\ldots,x_n,\alpha) = x_{(k_\ell)} \quad \text{und} \quad \mu_r = \mu_r(x_1,\ldots,x_n,\alpha) = x_{(k_r+1)}\,.$$

Da bei den geordneten Beobachtungen $x_{(i)}$ der Index *i* zwischen 1 und *n* liegen muß, ist unsere Definition der Vertrauensgrenzen nur sinnvoll, falls $k_\ell > 0$ und $k_r < n$, d.h. falls $(1-q)^n \le \alpha$ und $q^n \le \alpha$. Wenn *n* zu klein ist, versagt unsere Methode; wir setzen dann $\mu_\ell = -\infty$ bzw. $\mu_r = \infty$.

Es gilt nun die klassische Konfidenzinterpretation:

1) $P_F\left\{\mu_\ell(X_1,\ldots,X_n,\alpha) \le x_q\right\} \ge 1-\alpha$

2) $P_F\left\{x_q \le \mu_r(X_1,\ldots,X_n,\alpha)\right\} \ge 1-\alpha$

3) $P_F\left\{\mu_\ell(X_1,\ldots,X_n,\alpha) \le x_q \le \mu_r(X_1,\ldots,X_n,\alpha)\right\} \ge 1-2\alpha$

für jede beliebige unbekannte Verteilungsfunktion *F*, d.h. $\mu_\ell$ und $\mu_r$ sind Vertrauensgrenzen für $x_q$ zum Sicherheitsgrad $1-\alpha$ (bzw. $1-2\alpha$). Dabei sind keinerlei Voraussetzungen über die Verteilung *F* notwendig, auch keine Stetigkeit. Der Beweis ist analog wie im Fall des Medians.

## 4. Randomisierte Vertrauensgrenzen

Mit Hilfe von randomisierten Vertrauensgrenzen ist es möglich, bei stetiger Verteilungsfunktion *F* einen vorgegebenen Sicherheitsgrad $1-\alpha$ bzw. $1-2\alpha$ stets voll auszuschöpfen, da die randomisierten Vertrauensgrenzen im allgemeinen enger sind als die nichtrandomisierten. Bei größeren Stichproben (etwa im Beispiel am Ende dieses Kapitels) ist der Unterschied vernachlässigbar, und die Resultate dieses Abschnitts sind daher wohl eher von theoretischer Bedeutung. Es sei

$Y \sim Bi(n,\frac{1}{2})$, $U \sim U(0,1)$, $Y,U$ unabhängig;

$\tilde{Y} = Y + U$;

$(k_\ell,\delta_\ell)$, $(k_r,\delta_r)$ linkes und rechtes $\alpha$-Quantil zu $Bi(n,\frac{1}{2})$, d.h.

$$P\left\{Bi(n,\tfrac{1}{2}) < k_\ell\right\} + \delta_\ell P\left\{Bi(n,\tfrac{1}{2}) = k_\ell\right\} = \alpha$$

$$P\left\{Bi(n,\tfrac{1}{2}) > k_r\right\} + \delta_r P\left\{Bi(n,\tfrac{1}{2}) = k_r\right\} = \alpha.$$

Die Zufallsgröße $\tilde{Y} = Y + U$ ist stetig mit einer histogrammförmigen Dichte; die Punktwahrscheinlichkeit $P\{Y=k\}$ der Binomialverteilung $Bi(n,\frac{1}{2})$ wird ersetzt durch die zugehörige Gleichverteilung auf dem Intervall $[k,k+1)$. Wir bezeichnen daher die Verteilung von *Y* mit $\widetilde{Bi}(n,\frac{1}{2})$ (= stetige Variante der entsprechenden Binomialverteilung). Aus Symmetriegründen gilt: $k_\ell + k_r = n$ und $\delta_\ell = \delta_r$ $(= \delta)$. Weiter sei

$$\tilde{k}_\ell = k_\ell + \delta_\ell \quad \text{und} \quad \tilde{k}_r = k_r + 1 - \delta_r\,.$$

Dann gilt

$$P\left\{\tilde{Y} < \tilde{k}_\ell\right\} = P\left\{\tilde{Y} > \tilde{k}_r\right\} = \alpha\,,$$

d.h. $\tilde{k}_\ell$ und $\tilde{k}_r$ sind die $\alpha$-Quantile zu $\widetilde{Bi}(n,\frac{1}{2})$.

Nun seien $X_1,\ldots,X_n,U$ unabhängige Zufallsgrößen, die $X_i$ je mit der unbekannten Verteilungsfunktion *F*, *U* mit einer Gleichverteilung $U(0,1)$, und $\mu = med(F)$. Wir definieren die randomisierten Vertrauensgrenzen als

$$\tilde{\mu}_\ell = \tilde{\mu}_\ell(X_1,\ldots,X_n,U,\alpha) = X_{(\tilde{k}_\ell)} = \begin{cases} X_{(k_\ell)} & \text{falls } U > \delta \text{ (d.h. mit Wahrscheinlichkeit } 1-\delta) \\ X_{(k_\ell+1)} & \text{falls } U \le \delta \text{ (d.h. mit Wahrscheinlichkeit } \delta) \end{cases}$$

$$\tilde{\mu}_r = \tilde{\mu}_r(X_1,\ldots,X_n,U,\alpha) = X_{(\tilde{k}_r)} = \begin{cases} X_{(k_r+1)} & \text{falls } U > \delta \text{ (d.h. mit Wahrscheinlichkeit } 1-\delta) \\ X_{(k_r)} & \text{falls } U \le \delta \text{ (d.h. mit Wahrscheinlichkeit } \delta). \end{cases}$$

Man beachte, dass $k_\ell + k_r = n$, jedoch $\tilde{k}_\ell + \tilde{k}_r = n+1$; für $U > \delta$ ist $\tilde{\mu}_\ell = X_{(k_\ell)}$ und $\tilde{\mu}_r = X_{(k_r+1)}$ wie im nichtrandomisierten Fall, für $U \le \delta$ ist jedoch $\tilde{\mu}_\ell = X_{(k_\ell+1)}$ und $\tilde{\mu}_r = X_{(k_r)}$. Die Randomisierung führt also mit Wahrscheinlichkeit $\delta$ zu engeren Vertrauensgrenzen als im nichtrandomisierten Fall. Auch für die randomisierten Vertrauensgrenzen gilt die Konfidenzinterpretation, und falls die Verteilungsfunktion *F* stetig ist, so wird der Sicherheitsgrad voll ausgeschöpft, d.h. es gilt dann:

1) $P_F\{\tilde{\mu}_\ell(X_1,\ldots,X_n,U,\alpha) \le \mu\} = 1-\alpha$;
2) $P_F\{\mu \le \tilde{\mu}_r(X_1,\ldots,X_n,U,\alpha)\} = 1-\alpha$;
3) $P_F\{\tilde{\mu}_\ell(X_1,\ldots,X_n,U,\alpha) \le \mu \le \tilde{\mu}_r(X_1,\ldots,X_n,U,\alpha)\} = 1-2\alpha$.

*Beweis:*

Wenn die Verteilungsfunktion *F* stetig ist, können wir davon ausgehen, daß alle Beobachtungen $X_1,\ldots,X_n$ verschieden sind. Es sei

$$N_- = \#\{i \mid X_i < \mu\} = \text{ Anzahl der Beobachtungen } X_i \text{ mit } X_i < \mu, i = 1,\ldots,n.$$

Da *F* stetig ist, so gilt $P\{X_i < \mu\} = \frac{1}{2}$, und daher ist $N_- \sim Bi(n,\frac{1}{2})$.

Es sei $\tilde{N}_- = N_- + U$, dann gilt $\tilde{N}_- \sim \widetilde{Bi}(n,\frac{1}{2})$. Nun sei $\tilde{\mu}_\ell = \tilde{\mu}_\ell(X_1,\ldots,X_n,\alpha)$ definiert wie oben. Dann gilt

$$\{\mu < \tilde{\mu}_\ell\} = \{\mu < X_{(\tilde{k}_\ell)}\} = \{\tilde{N}_- < \tilde{k}_\ell\},$$

und daher gilt

$$P\{\mu < \tilde{\mu}_\ell\} = P\{\mu < X_{(\tilde{k}_\ell)}\} = P\{\tilde{N}_- < \tilde{k}_\ell\} = P\{\widetilde{Bi}(n,\tfrac{1}{2}) < \tilde{k}_\ell\} = \alpha.$$

Daraus ergibt sich

$$P\{\tilde{\mu}_\ell \le \mu\} = 1-\alpha,$$

und damit ist die Behauptung 1) bewiesen. Der Beweis der Behauptungen 2) und 3) ist analog.

*Mediantreue Punktschätzung*

Wir betrachten hier den Spezialfall $\alpha = \frac{1}{2}$. Dann fallen die beiden randomisierten Vertrauensgrenzen zusammen, und wir erhalten eine sogenannte mediantreue Punktschätzung für $\mu$.

Falls *n* ungerade ist, so ist $k_\ell = k_r + 1 = \frac{1}{2}(n+1)$ und $\delta_\ell = \delta_r = 0$. Die Randomisierung fällt also weg, und es ist $\tilde{\mu}_\ell = \tilde{\mu}_r = \hat{\mu} = X_{(k_\ell)} =$ Zentralwert von $X_1,\ldots,X_n$. Für die Punktschätzung $\hat{\mu}$ gilt aufgrund der Konfidenzeigenschaft $P\{\hat{\mu} < \mu\} = P\{\hat{\mu} > \mu\} = \frac{1}{2}$, und daher ist $\hat{\mu}$ eine mediantreue Schätzung für $\mu = med(F)$ bei beliebiger Verteilungsfunktion *F* (wie wir bereits oben gesehen haben).

Falls *n* gerade ist, so ist $k_\ell = k_r = n/2$ und $\delta_\ell = \delta_r = \frac{1}{2}$. Daher ist dann

$$\tilde{\mu}_\ell = \tilde{\mu}_r = \hat{\mu} = \begin{cases} X_{(k_\ell)} & \text{falls } U > \frac{1}{2} \text{ (d.h. mit Wahrscheinlichkeit } \frac{1}{2}) \\ X_{(k_{\ell+1})} & \text{falls } U \le \frac{1}{2} \text{ (d.h. mit Wahrscheinlichkeit } \frac{1}{2}). \end{cases}$$

Die randomisierte Punktschätzung $\hat{\mu}$ ist wiederum aufgrund der Konfidenzeigenschaft mediantreu für $\mu = med(F)$. Der Zentralwert von $X_1,\ldots,X_n$ wird bei geradem *n* üblicherweise definiert als $Ze = \frac{1}{2}(X_{(k_\ell)} + X_{(k_\ell+1)})$. Man beachte, daß der randomisierte Schätzer $\hat{\mu}$ nicht identisch ist zum Zentralwert; der Zentralwert hängt nur ab von den Beobachtungen $X_1,\ldots,X_n$, während $\hat{\mu}$ auch noch vom Wert der Hilfsvariablen *U* abhängt. Das folgende Beispiel zeigt, daß der Zentralwert (bei geradem *n*) i.a. kein mediantreuer Schätzer für den Median $\mu = med(F)$ ist.

*Beispiel mit* $n = 2$:

$X_1, X_2$ seien unabhängige Zufallsgrößen je mit der Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ x^2 & \text{für } 0 \le x \le 1 \\ 1 & \text{für } x > 1. \end{cases}$$

Dann ist $\mu = \text{med}(F) = \frac{1}{2}\sqrt{2} = 0.7071$. Der Zentralwert $Y = \frac{1}{2}(X_1 + X_2)$ besitzt die Verteilungsfunktion

$$G(y) = \begin{cases} \frac{8}{3}y^4 & \text{für } 0 \le y \le \frac{1}{2} \\ \frac{1}{3}\left(-8y^4 + 24y^2 - 16y + 3\right) & \text{für } \frac{1}{2} < y \le 1, \end{cases}$$

und der Median der Verteilungsfunktion *G* ist $\mu_G = med(G) = 0.6789 \neq med(F)$. Der randomisierte Schätzer $\hat{\mu}$ ist also mediantreu für $\mu = med(F)$, der Zentralwert *Y* jedoch nicht.

## 5. Ein-Stichproben-Problem für den Median

$X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit der unbekannten Verteilungsfunktion *F*;
$\mu = med(F)$ = Median von *F*;
$\mu_0 =$ theoretischer Vergleichsparameter.

*Testprobleme:*

1) $H_0 : \mu \le \mu_0$ gegen $H_1 : \mu > \mu_0$ (einseitiges Testproblem)
2) $H_0 : \mu \ge \mu_0$ gegen $H_1 : \mu < \mu_0$ (einseitiges Testproblem)
3) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ (zweiseitiges Testproblem)

$x_1,\ldots,x_n$ Realisationen zu $X_1,\ldots,X_n$.

Sind die beobachteten Realisationen $x_1,\ldots,x_n$ vereinbar mit $H_0$, oder muß $H_0$ zugunsten von $H_1$ abgelehnt werden?

*Lösung:*

Bestimme die linke und rechte Vertrauensgrenze für den Median $\mu$:

$$\mu_\ell = \mu_\ell(x_1,\ldots,x_n,\alpha) = x_{(k_\ell)} \quad \text{und} \quad \mu_r = \mu_r(x_1,\ldots,x_n,\alpha) = x_{(k_r+1)}.$$

1) Verwirf $H_0$, falls $\mu_\ell > \mu_0$. Dann gilt: Irrtumswahrscheinlichkeit 1. Art $\le \alpha$.
2) Verwirf $H_0$, falls $\mu_r < \mu_0$. Dann gilt: Irrtumswahrscheinlichkeit 1. Art $\le \alpha$.
3) Verwirf $H_0$, falls $\mu \notin [\mu_\ell, \mu_r]$. Dann gilt: Irrtumswahrscheinlichkeit 1. Art $\le 2\alpha$.

*Beweis von 1):*

Falls $\mu \le \mu_0$ (d.h. $H_0$ ist wahr) und $\mu_\ell = \mu_\ell(X_1,\ldots,X_n,\alpha)$, so gilt $\{\mu_0 < \mu_\ell\} \subset \{\mu < \mu_\ell\}$, und wir erhalten wegen der Konfidenzeigenschaft von $\mu_\ell = \mu_\ell(X_1,\ldots,X_n,\alpha)$

$$P_F\{\text{Test verwirft } H_0\} = P_F\{\mu_0 < \mu_\ell\} \le P_F\{\mu < \mu_\ell\} \le \alpha,$$

und damit ist die Behauptung 1) bewiesen. Der Beweis von 2) und 3) ist analog.

## 6. Zwei-Stichproben-Problem für den Median

$\left.\begin{matrix} X_1,\ldots,X_{n_1} \\ Y_1,\ldots,Y_{n_2} \end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases} \text{je mit der V.fkt. } F \\ \text{je mit der V.fkt. } G \end{cases}$

$\mu_F = med(F) =$ Median von $F$

$\mu_G = med(G) =$ Median von $G$

*Testprobleme:*

1) $H_0$: $\mu_F \le \mu_G$ gegen $H_0$: $\mu_F > \mu_G$ (einseitiges Testproblem)
2) $H_0$: $\mu_F \ge \mu_G$ gegen $H_0$: $\mu_F < \mu_G$ (einseitiges Testproblem)
3) $H_0$: $\mu_F = \mu_G$ gegen $H_0$: $\mu_F \ne \mu_G$ (zweiseitiges Testproblem)

$x_1,\ldots,x_{n_1}$ Realisationen zu $X_1,\ldots,X_{n_1}$

$y_1,\ldots,y_{n_2}$ Realisationen zu $Y_1,\ldots,Y_{n_2}$

Sind die beobachteten Realisationen $x_1,\ldots,x_{n_1}$ und $y_1,\ldots,y_{n_2}$ vereinbar mit $H_0$, oder muß $H_0$ zugunsten von $H_1$ abgelehnt werden?

*Lösung:*

Bestimme zu den beiden Stichproben $\mathbf{x} = (x_1,\ldots,x_{n_1})$ und $\mathbf{y} = (y_1,\ldots,y_{n_2})$ die Vertrauensgrenzen für den Median $\mu_F$ bzw. $\mu_G$

$\mu_\ell(\mathbf{x},\alpha) =$ linke Vertrauensgrenze für $\mu_F$ aufgrund von $\mathbf{x} = (x_1,\ldots,x_{n_1})$
$\mu_r(\mathbf{x},\alpha) =$ rechte Vertrauensgrenze für $\mu_F$ aufgrund von $\mathbf{x} = (x_1,\ldots,x_{n_1})$

$\mu_\ell(\mathbf{y},\alpha) =$ linke Vertrauensgrenze für $\mu_G$ aufgrund von $\mathbf{y} = (y_1,\ldots,y_{n_2})$
$\mu_r(\mathbf{y},\alpha) =$ rechte Vertrauensgrenze für $\mu_G$ aufgrund von $\mathbf{y} = (y_1,\ldots,y_{n_2})$

1) Verwirf $H_0$, falls $\mu_\ell(\mathbf{x},\alpha) > \mu_r(\mathbf{y},\alpha)$. Dann gilt: Irrtumswahrscheinlichkeit 1. Art $\le 2\alpha$;
2) Verwirf $H_0$, falls $\mu_r(\mathbf{x},\alpha) < \mu_\ell(\mathbf{y},\alpha)$. Dann gilt: Irrtumswahrscheinlichkeit 1. Art $\le 2\alpha$;
3) Verwirf $H_0$, falls die beiden Intervalle $[\mu_\ell(\mathbf{x},\alpha), \mu_r(\mathbf{x},\alpha)]$ und $[\mu_\ell(\mathbf{y},\alpha), \mu_r(\mathbf{y},\alpha)]$ disjunkt sind. Dann gilt: Irrtumswahrscheinlichkeit 1. Art $\le 4\alpha$.

*Beweis von 1):*

Es sei $\mu_F \le \mu_G$, d.h. $H_0$ sei wahr. Wir betrachten die beiden zufälligen Ereignisse

$$A = \left\{(\mathbf{x},\mathbf{y}) \middle| \mu_\ell(\mathbf{x},\alpha) \le \mu_F, \mu_G \le \mu_r(\mathbf{y},\alpha)\right\}$$

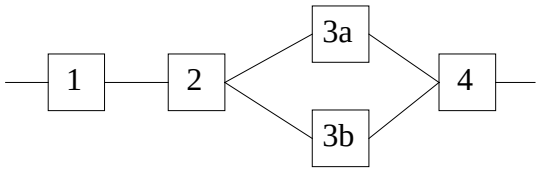
$$B = \left\{(\mathbf{x},\mathbf{y}) \middle| \mu_\ell(\mathbf{x},\alpha) \le \mu_r(\mathbf{y},\alpha)\right\}.$$

Dann gilt $P(A) \ge (1-\alpha)^2 > 1-2\alpha$, und da $A \subset B$, so ist $B^c \subset A^c$, und wir erhalten

$$P\left(B^c\right) \le P\left(A^c\right) \le 2\alpha,$$

und damit ist die Behauptung 1) bewiesen. Der Beweis von 2) und 3) ist analog.

## 7. Simulationsexperiment zur Lebensdauer eines Gerätes




Wir stellen uns die Aufgabe, die mittlere Lebensdauer eines Gerätes (vgl. nebenstehende Abbildung) abzuschätzen, wenn die Wahrscheinlichkeitsverteilung der Lebensdauer der fünf einzelnen Bauelemente bekannt ist. Wir bezeichnen die entsprechenden Zufallsgrößen mit $T_1$, $T_2$, $T_{3a}$, $T_{3b}$, $T_4$. Wir nehmen an, daß die fünf Zufallsgrößen unabhängig voneinander sind und je eine Exponentialverteilung besitzen mit der Dichte $f(t) = e^{-t}$, $t \ge 0$. besitzen. Die Lebensdauer des Gerätes ist dann gegeben durch

$$T = \min(T_1, T_2, T_3, T_4), \text{ wobei } T_3 = \max(T_{3a}, T_{3b}).$$

In einem Simualtionsexperiment wollen wir $M = 10000$ Geräte bauen, die Lebensdauer dieser Geräte bestimmen, und aufgrund der $M = 10000$ Realisationen von *T* Vertrauensgrenzen für den Erwartungswert, den Median und das untere Quartil der Verteilung von *T* bestimmen. (Problem aus: I. M. Sobol, Die Monte-Carlo Methode, Deutscher Verlag der Wissenschaften, Berlin 1991).

*Lösung mit Minitab:*

1. Wir setzen für den Zufallszahlengenerator den Startwert 77 (damit das Experiment wiederholbar wird).
2. Wir erzeugen je $M = 10\,000$ Realisationen für die Komponenten $T_1, T_2, T_{3a}, T_{3b}, T_4$. Jede Zeile des Arbeitsblatts entspricht einem einzelnen Gerät, die $M = 10\,000$ Geräte sind nun gebaut (vgl. Tabelle 1).
3. Wir bestimmen die Lebensdauer *T* der einzelnen Geräte gemäß der Formel
   $$T = \min(T_1, T_2, T_3, T_4), \text{ wobei } T_3 = \max(T_{3a}, T_{3b}).$$
   Die Spalte (8) von Tabelle 1 enthält nun die $M = 10\,000$ Realisationen $t_1, \ldots, t_M$ der Zufallsgröße *T*, während die Spalte (9) die geordneten Beobachtungen $t_{(1)}, \ldots, t_{(M)}$ enthält.

*Klassische Punktschätzung und Vertrauensgrenzen* für den Erwartungswert $\mu$ der Zufallsgröße *T*:

$1-2\alpha = 0.99 =$ Sicherheitsgrad, d.h. $\alpha = 0.005$, $z_{1-\alpha} = 2.576$

$\overline{t} = 0.3001 =$ arithmetisches Mittel von $t_1,\ldots,t_M$

$s = 0.2830 =$ empirische Standardabweichung von $t_1,\ldots,t_M$ $(s^2 \approx \sigma^2 = Var(T),$ da $M$ groß)

$\hat{\mu} = \overline{t} = 0.3001 =$ Punktschätzung für $\mu$

$$\mu_\ell = \overline{t} - z_{1-\alpha}\frac{\sigma}{\sqrt{M}} \approx 0.3001 - 2.576\frac{0.2830}{100} = 0.3001 - 0.0073 = 0.2928$$

$$\mu_r = \overline{t} + z_{1-\alpha}\frac{\sigma}{\sqrt{M}} \approx 0.3001 + 0.0073 = 0.3074$$

Mit einem Sicherheitsgrad von 99% liegt der Erwartungswert $\mu$ der Verteilung von *T* zwischen $\mu_\ell = 0.2928$ und $\mu_r = 0.3074$; Länge des Vertrauensintervalls: $\mu_r - \mu_\ell = 0.0146$.

*Verteilungsfreie Vertrauensgrenzen für den Median* $\mu$ der Zufallsgröße *T:*

$1-2\alpha = 0.99 =$ Sicherheitsgrad, d.h. $\alpha = 0.005$

$k_\ell =$ linke kritische $\alpha$-Schranke zu $Bi(M,\frac{1}{2}) = 4871$

$k_r =$ rechte kritische $\alpha$-Schranke zu $Bi(M,\frac{1}{2}) = 5129$

$\hat{\mu} =$ Zentralwert $= \frac{1}{2}(t_{(5000)} + t_{(5001)}) = 0.2159 =$ Punktschätzung für Median $\mu$

$\mu_\ell = t_{(k_\ell)} = t_{(4871)} = 0.2085$

$\mu_r = t_{(k_r+1)} = t_{(5130)} = 0.2236$

Mit einem Sicherheitsgrad von 99% liegt der Median $\mu$ der Verteilung von *T* zwischen $\mu_\ell = 0.2085$ und $\mu_r = 0.2236$; Länge des Vertrauensintervalls: $\mu_r - \mu_\ell = 0.0151$.

Obschon wir hier keine Annahme über die Verteilung von *T* machen, ist das Vertrauensintervall für den Median praktisch gleich lang wie das klassische Vertrauensintervall für den Erwartungswert; beim klassischen Verfahren setzen wir voraus, dass das arithmetische Mittel $\overline{t}$ (aufgefasst als Zufallsgröße) aufgrund des zentralen Grenzwertsatzes nahezu normalverteilt ist.

*Verteilungsfreie Vertrauensgrenzen für das untere Quartil* $t_{0.25}$ der Zufallsgröße *T:*

$1-2\alpha = 0.99 =$ Sicherheitsgrad, d.h. $\alpha = 0.005$

$k_\ell =$ linke kritische $\alpha$-Schranke zu $Bi(M,\frac{1}{4}) = 2389$

$k_r =$ rechte kritische $\alpha$-Schranke zu $Bi(M,\frac{1}{4}) = 2612$

$\hat{t}_{0.25} = Q_1 =$ unteres empirisches Quartil $\approx x_{(2500)} = 0.0952 =$ Punktschätzung für $t_{0.25}$

$t_\ell = t_{(k_\ell)} = t_{(2389)} = 0.0906$

$t_r = t_{(k_r+1)} = t_{(2613)} = 0.0994$

Mit einem Sicherheitsgrad von 99% liegt das untere Quartil $t_{0.25}$ der Verteilung von *T* zwischen $t_\ell = 0.0906$ und $t_r = 0.0994$. Dabei sind bezüglich der Verteilung der Zufallsgröße *T* keinerlei Annahmen erforderlich.

**Tabelle 1:** Lebensdauer der $M = 10\,000$ Komponenten

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|
| $i$ | $T_1$ | $T_2$ | $T_{3a}$ | $T_{3b}$ | $T_3$ | $T_4$ | $T$ | $T$ (geordn.) |
| 1 | 0.4452 | 1.5022 | 2.1116 | 0.6802 | 2.1116 | 1.0412 | 0.4452 | 0.0000 |
| 2 | 0.1230 | 2.8175 | 0.8093 | 0.6457 | 0.8093 | 0.3240 | 0.1230 | 0.0000 |
| 3 | 0.8048 | 0.7292 | 1.9312 | 0.3040 | 1.9312 | 0.8647 | 0.7292 | 0.0001 |
| 4 | 0.7570 | 1.8175 | 0.3836 | 0.8890 | 0.8890 | 0.2074 | 0.2074 | 0.0001 |
| 5 | 0.0633 | 0.8056 | 1.9822 | 2.6324 | 2.6324 | 1.3095 | 0.0633 | 0.0001 |
| 6 | 0.0850 | 0.1954 | 0.2907 | 0.2039 | 0.2907 | 0.0599 | 0.0599 | 0.0001 |
| 7 | 1.8681 | 0.2001 | 2.9640 | 1.6481 | 2.9640 | 0.0614 | 0.0614 | 0.0002 |
| 8 | 7.7553 | 3.0531 | 1.0570 | 0.7549 | 1.0570 | 0.6724 | 0.6724 | 0.0002 |
| 9 | 0.1439 | 0.8307 | 1.5564 | 0.1723 | 1.5564 | 0.3681 | 0.1439 | 0.0002 |
| 10 | 1.4811 | 0.5753 | 0.4147 | 2.2989 | 2.2989 | 0.5400 | 0.5400 | 0.0003 |
| … | … | … | … | … | … | … | … | … |
| 4865 | 0.8872 | 0.0891 | 0.0494 | 1.4410 | 1.4410 | 0.0674 | 0.0674 | 0.2081 |
| 4866 | 0.6695 | 0.3222 | 0.6869 | 0.1637 | 0.6869 | 1.5083 | 0.3222 | 0.2083 |
| 4867 | 0.6812 | 1.0694 | 0.0236 | 1.0711 | 1.0711 | 0.7919 | 0.6812 | 0.2083 |
| 4868 | 3.1483 | 0.9609 | 0.1047 | 0.0817 | 0.1047 | 2.2268 | 0.1047 | 0.2084 |
| 4869 | 0.2063 | 0.3222 | 1.9140 | 0.2212 | 1.9140 | 0.3681 | 0.2063 | 0.2084 |
| 4870 | 0.1469 | 0.4914 | 0.0434 | 0.5579 | 0.5579 | 0.2765 | 0.1469 | 0.2085 |
| 4871 | 0.1525 | 1.7824 | 0.3301 | 2.1982 | 2.1982 | 1.2149 | 0.1525 | 0.2085 |
| 4872 | 0.6384 | 0.0981 | 0.8310 | 0.5277 | 0.8310 | 2.3854 | 0.0981 | 0.2086 |
| 4873 | 0.9505 | 0.4512 | 0.8294 | 2.5847 | 2.5847 | 1.5464 | 0.4512 | 0.2088 |
| 4874 | 0.0971 | 0.1909 | 0.0753 | 1.0860 | 1.0860 | 1.9543 | 0.0971 | 0.2088 |
| 4875 | 0.2754 | 1.9035 | 0.3951 | 3.4493 | 3.4493 | 0.2481 | 0.2481 | 0.2088 |
| … | … | … | … | … | … | … | … | … |
| 4995 | 1.3604 | 0.4533 | 1.1656 | 0.0515 | 1.1656 | 0.1224 | 0.1224 | 0.2157 |
| 4996 | 2.2505 | 0.6697 | 1.6563 | 1.1180 | 1.6563 | 0.6526 | 0.6526 | 0.2157 |
| 4997 | 0.8182 | 3.3312 | 0.3782 | 0.8014 | 0.8014 | 0.3674 | 0.3674 | 0.2158 |
| 4998 | 0.0824 | 0.4048 | 0.9661 | 2.7316 | 2.7316 | 0.0099 | 0.0099 | 0.2158 |
| 4999 | 0.7971 | 1.5699 | 2.4555 | 1.1110 | 2.4555 | 0.6958 | 0.6958 | 0.2159 |
| 5000 | 2.5417 | 2.8150 | 0.2699 | 1.5328 | 1.5328 | 0.4061 | 0.4061 | 0.2159 |
| 5001 | 0.9504 | 1.6352 | 0.3547 | 0.0778 | 0.3547 | 0.1576 | 0.1576 | 0.2159 |
| 5002 | 1.0679 | 0.3708 | 0.7666 | 1.5616 | 1.5616 | 0.2069 | 0.2069 | 0.2161 |
| 5003 | 0.7067 | 0.8135 | 1.6292 | 0.3977 | 1.6292 | 1.6295 | 0.7067 | 0.2161 |
| 5004 | 0.0901 | 0.0763 | 0.1361 | 0.4309 | 0.4309 | 0.1870 | 0.0763 | 0.2162 |
| 5005 | 1.5863 | 0.7736 | 1.6666 | 1.9117 | 1.9117 | 0.2636 | 0.2636 | 0.2163 |
| … | … | … | … | … | … | … | … | … |
| 5125 | 0.9014 | 0.1115 | 0.7495 | 0.6856 | 0.7495 | 2.3975 | 0.1115 | 0.2234 |
| 5126 | 0.3352 | 1.5197 | 0.6408 | 0.3671 | 0.6408 | 0.5032 | 0.3352 | 0.2235 |
| 5127 | 0.2591 | 1.8623 | 1.4195 | 1.8108 | 1.8108 | 0.2398 | 0.2398 | 0.2236 |
| 5128 | 1.3165 | 1.5202 | 1.7193 | 0.8044 | 1.7193 | 0.9715 | 0.9715 | 0.2236 |
| 5129 | 0.6629 | 0.2745 | 0.6989 | 0.0404 | 0.6989 | 0.2804 | 0.2745 | 0.2236 |
| 5130 | 0.8120 | 3.5162 | 0.5221 | 2.8928 | 2.8928 | 1.2786 | 0.8120 | 0.2236 |
| 5131 | 0.2096 | 0.6615 | 0.2254 | 1.0011 | 1.0011 | 0.9850 | 0.2096 | 0.2237 |
| 5132 | 0.8450 | 0.2017 | 0.3206 | 0.4710 | 0.4710 | 0.0908 | 0.0908 | 0.2239 |
| 5133 | 0.0997 | 0.1251 | 0.0908 | 3.0355 | 3.0355 | 1.0482 | 0.0997 | 0.2239 |
| 5134 | 0.0335 | 1.3991 | 2.6597 | 0.0929 | 2.6597 | 1.3050 | 0.0335 | 0.2240 |
| 5135 | 2.0280 | 0.6007 | 1.8601 | 0.0711 | 1.8601 | 0.5426 | 0.5426 | 0.2240 |
| … | … | … | … | … | … | … | … | … |
| 9990 | 0.4722 | 0.8589 | 1.2043 | 0.1574 | 1.2043 | 0.5618 | 0.4722 | 1.8496 |
| 9991 | 1.4736 | 0.2698 | 0.4790 | 0.4285 | 0.4790 | 0.8474 | 0.2698 | 1.8647 |
| 9992 | 0.0947 | 0.5282 | 1.8048 | 1.5016 | 1.8048 | 1.8918 | 0.0947 | 1.8661 |
| 9993 | 3.9177 | 3.4234 | 0.0795 | 1.8935 | 1.8935 | 0.2046 | 0.2046 | 1.9141 |
| 9994 | 0.4363 | 5.0523 | 2.0013 | 0.4488 | 2.0013 | 0.1969 | 0.1969 | 1.9530 |
| 9995 | 0.2772 | 1.7499 | 1.5158 | 0.2613 | 1.5158 | 1.3607 | 0.2772 | 2.0215 |
| 9996 | 0.9345 | 0.2464 | 0.4228 | 1.1977 | 1.1977 | 3.1602 | 0.2464 | 2.0769 |
| 9997 | 0.5954 | 4.0259 | 1.3269 | 0.4285 | 1.3269 | 0.1687 | 0.1687 | 2.1540 |
| 9998 | 0.7609 | 0.5323 | 0.7683 | 1.0333 | 1.0333 | 0.6748 | 0.5323 | 2.3706 |
| 9999 | 1.5825 | 1.9019 | 0.3786 | 3.1524 | 3.1524 | 0.0643 | 0.0643 | 2.5592 |
| 10000 | 0.4500 | 0.2536 | 0.8699 | 1.8436 | 1.8436 | 1.2387 | 0.2536 | 2.5685 |

# 4. Vorzeichentest

## 1. Vorzeichentest

Wir haben den Vorzeichentest (Zeichentest) bereits im Kapitel 2 kennengelernt. Da wir dort den Zeichentest mit dem klassischen t-Test vergleichen wollten, haben wir uns auf den stetigen Fall beschränkt. Hier wollen wir den Zeichentest im allgemeinen Fall untersuchen. Es seien

$X_1, \cdots, X_n$ unabhängige Zufallsgrößen je mit der gleichen Verteilung:

$P\{X_i > 0\} = p_+$ (positive Wirkung);

$P\{X_i < 0\} = p_-$ (negative Wirkung);

$P\{X_i = 0\} = p_0$ (keine Wirkung);

$(p_+ + p_- + p_0 = 1)$.

Wir betrachten das Testproblem

$H_0 : p_+ \leq p_-$ gegen $H_1 : p_+ > p_-$.

Die Alternative besagt also, dass die Wahrscheinlichkeit für eine positive Wirkung größer ist als die Wahrscheinlichkeit für eine negative Wirkung, dass das Medikament also insgesamt eine positive Wirkung besitzt.

*Bemerkung zum Testproblem*

Es sei $F$ die Verteilungsfunktion der Zufallsgrößen $X_i$ und $\mu = med(F)$. Falls $F$ stetig ist, und der Median $\mu$ eindeutig bestimmt ist, so gilt offensichtlich

$p_+ = p_- \Leftrightarrow \mu = 0$

$p_+ < p_- \Leftrightarrow \mu < 0$

$p_+ > p_- \Leftrightarrow \mu > 0$;

somit kann dann unser obiges Testproblem auch geschrieben werden als $H_0 : \mu \leq 0$ gegen $H_1 : \mu > 0$; unser Testproblem ist dann äquivalent zum Einstichprobenproblem für den Median. Dies ist jedoch nicht mehr richtig im Fall $p_0 > 0$ wie das folgende Beispiel zeigt. Es sei $p_0 = \frac{1}{2}$, $p_- = \frac{1}{6}$ und $p_+ = \frac{2}{6}$. Dann ist offensichtlich $p_+ > p_-$, aber es gilt $\mu = med(F) = 0$, denn $P\{X < 0\} = p_- < \frac{1}{2}$ und $P\{X > 0\} = p_+ < \frac{1}{2}$. Das Einstichprobenproblem für den Median ist also nicht äquivalent zum Testproblem beim Vorzeichentest.

Nun sei

$N_+ = \#\{i \mid X_i = +1\}$;

$N_- = \#\{i \mid X_i = -1\}$;

$N_0 = \#\{i \mid X_i = 0\}$.

Dann gilt:

$(N_+, N_-, N_0) \sim \mathrm{Mu}(n, p_+, p_-, p_0)$.

*Vorzeichentest (Variante A):*

Die $N_0$ Beobachtungen mit $X_i = 0$ werden zufällig nach + oder – zugeteilt: $N_0 = N_{0+} + N_{0-}$; dann gilt:

$N_{0+} \sim Bi(n, \frac{1}{2} p_0)$, $N_{0-} \sim Bi(n, \frac{1}{2} p_0)$;

$\tilde{N}_+ = N_+ + N_{0+} \sim Bi(n, \tilde{p}_+)$ mit $\tilde{p}_+ = p_+ + \frac{1}{2} p_0$;

$\tilde{N}_- = N_- + N_{0-} \sim Bi(n, \tilde{p}_-)$ mit $\tilde{p}_- = p_- + \frac{1}{2} p_0$.

Da $\tilde{p}_+ + \tilde{p}_- = p_+ + p_- + p_0 = 1$, so kann das Testproblem nun geschrieben werden als:

$H_0 : p_+ \le p_- \Leftrightarrow \tilde{p}_+ \le \tilde{p}_- \Leftrightarrow \tilde{p}_+ \le \frac{1}{2}$;

$H_1 : p_+ > p_- \Leftrightarrow \tilde{p}_+ > \tilde{p}_- \Leftrightarrow \tilde{p}_+ > \frac{1}{2}$.

Bestimme $k_r = k_r(n, \alpha) =$ rechte kritische $\alpha$-Schranke zu $Bi\left(n, \frac{1}{2}\right)$;
verwirf $H_0$, falls $\tilde{N}_+ > k_r(n, \alpha)$.

Bei diesem Test handelt es sich also um den einfachen Binomialtest: $Y \sim Bi(n, p)$, $H_0$: $p \le \frac{1}{2}$ gegen $H_1$: $p > \frac{1}{2}$, wobei $Y = \tilde{N}_+$ und $p = \tilde{p}_+$. Der Test hält das Signifikanzniveau $\alpha$ ein, und er ist konsistent.

(Bei der einseitigen Fragestellung in die andere Richtung und bei der zweiseitigen Fragestellung ist alles analog.)

*Vorzeichentest (Variante B):*

$N_+ \,|\, N_+ + N_- = m \sim Bi(m, \vartheta)$, wobei $\vartheta = \dfrac{p_+}{p_+ + p_-}$;

$H_0 : p_+ \le p_- \Leftrightarrow \vartheta \le \frac{1}{2}$;

$H_1 : p_+ > p_- \Leftrightarrow \vartheta > \frac{1}{2}$.

Berechne $m = N_+ + N_-$.
Bestimme $k_r = k_r(m, \alpha) =$ rechte kritische $\alpha$-Schranke zu $Bi\left(m, \frac{1}{2}\right)$.

Verwirf $H_0$, falls $N_+ > k_r(m, \alpha)$.
Der Test hält das Signifikanzniveau $\alpha$ ein, und er ist konsistent.

(Bei der einseitigen Fragestellung in die andere Richtung und bei der zweiseitigen Fragestellung ist auch hier alles analog.)

Nun stellt sich die Frage, welcher der beiden Tests (Variante A oder B) der bessere ist.

## 2. Asymptotische Gütefunktion des Tests A

Bei zufälliger Aufteilung der Beobachtungen mit $X_i = 0$ nach + oder – gilt

$Y = N_+ + N_{0+} \sim Bi(n, p)$ mit $p = p_+ + \frac{1}{2} p_0$,

und das Testproblem lautet:

$H_0$: $p \le \frac{1}{2}$ gegen $H_1$: $p > \frac{1}{2}$.

Die Testvorschrift lautet: Verwirf $H_0$, falls $Y > k_r$, wobei $k_r = k_r(n, \alpha)$ die rechte kritische $\alpha$-Schranke ist zur symmetrischen Binomialverteilung $Bi(n, \frac{1}{2})$. Wir wollen die asymptotische Güte-

funktion dieses Tests bestimmen. Da der Test konsistent ist, haben wir wiederum (wie in Kapitel 2) die Auflösung zu verfeinern mit wachsendem Stichprobenumfang *n*. Wir führen also die folgende Parametertransformation durch:

$\delta = d\sqrt{n}$, wobei $d = p_+ - p_-$.

Das Testproblem kann jetzt geschrieben werden als

$H_0$: $d \le 0$ gegen $H_1$: $d > 0$

oder äquivalent dazu als

$H_0$: $\delta \le 0$ gegen $H_1$: $\delta > 0$.

Nun wollen wir den Grenzwert der Gütefunktion für $n \to \infty$ ($\delta$, $p_0$, $\alpha$ fest) bestimmen. Es sei

(1) $\beta_n(\delta) = P\{\text{Test A verwirft } H_0\} = P\{Y > k_r(n,\alpha)\} = P\left\{\frac{Y-\mu}{\sigma} > \frac{k_r-\mu}{\sigma}\right\}$,

wobei $\mu = E(Y) = np$ und $\sigma^2 = Var(Y) = np(1-p)$. Aufgrund des zentralen Grenzwertsatzes gilt bei unserem Grenzübergang

$$\frac{Y-\mu}{\sigma} \to N(0,1),$$

und es ist

$$\frac{k_r-\mu}{\sigma} = \left(\frac{k_r-\mu_0}{\sigma_0} - \frac{\mu-\mu_0}{\sigma_0}\right)\frac{\sigma_0}{\sigma},$$

wobei $\mu_0 = E_0(Y) = n/2$ und $\sigma_0^2 = Var_0(Y) = n/4$. Aufgrund des zentralen Grenzwertsatzes gilt

$$\frac{k_r-\mu_0}{\sigma_0} \longrightarrow z_{1-\alpha},$$

und weiter ist

$$p = p_+ + \tfrac{1}{2}p_0 = \tfrac{1}{2}(p_+ + p_- + p_0) + \tfrac{1}{2}(p_+ - p_-) = \tfrac{1}{2} + \tfrac{1}{2}d = \tfrac{1}{2} + \frac{\delta}{2\sqrt{n}}.$$

Somit erhalten wir bei unserem Grenzübergang $n \to \infty$ ($\delta$, $p_0$, $\alpha$ fest)

$$p \to \tfrac{1}{2},\ \frac{\sigma_0}{\sigma} = \frac{1}{2\sqrt{p(1-p)}} \to 1 \text{ und } \frac{\mu-\mu_0}{\sigma_0} = \frac{n(p-\frac{1}{2})}{\sqrt{n/4}} = 2\sqrt{n}\,\frac{\delta}{2\sqrt{n}} = \delta.$$

Damit gilt

$$\frac{k_r-\mu}{\sigma} \to z_{1-\alpha} - \delta,$$

und wir erhalten aus (1) beim Grenzübergang $n \to \infty$ ($\delta$, $p_0$, $\alpha$ fest)

(2) $\beta_n(\delta) \to \beta(\delta) = P\{N(0,1) > z_{1-\alpha} - \delta\} = P\{N(0,1) < -z_{1-\alpha} + \delta\}$,

und damit ist die gesuchte Grenzfunktion gefunden.

## 3. Asymptotische Gütefunktion des Tests B

Nun wollen wir die asymptotische Gütefunktion des Tests B bestimmen. Die Zufallsgrößen $N_+, N_-, N_0$ seien definiert wir in Abschnitt 1, und wir betrachten das Testproblem $H_0$: $p_+ \le p_-$ gegen $H_1$: $p_+ > p_-$. Test B verwirft die Nullhypothese, falls $N_+ > k_r$, wobei $k_r = k_r(m,\alpha) =$ rechte kritische $\alpha$-Schranke zu $Bi(m,\frac{1}{2})$ mit $m = N_+ + N_-$; die Beobachtungen mit $X_i = 0$ bleiben unberücksichtigt. Wie im Abschnitt 2 führen wir neue Parameter ein

$$\delta = d\sqrt{n}, \text{ wobei } d = p_+ - p_-,$$

und unser Testproblem lautet dann genau wie dort $H_0$: $\delta \le 0$ gegen $H_1$: $\delta > 0$. Wir wollen die asymptotische Gütefunktion von Test B bestimmen für $n \to \infty$ ($\delta$, $p_0$, $\alpha$ fest). Wir können schreiben

$$(3)\quad \beta_n(\delta) = P\{\text{Test B verwirft } H_0\} = \sum_{m=0}^{n} P\{N_0 = n-m\} \underbrace{P\{N_+ > k_r(m,\alpha) \,|\, N_0 = n-m\}}_{=\beta_{n|m}(\delta)}.$$

Beachte, dass $\{N_0 = n-m\} = \{N_+ + N_- = m\}$. Da

$$N_+ \,|\, N_0 = n-m \;\sim\; Bi(m,\vartheta) \text{ mit } \vartheta = \frac{p_+}{p_+ + p_-},$$

so kann die Gütefunktion des bedingten Tests geschrieben werden als

$$\beta_{n|m}(\delta) = P\{N_+ > k_r(m,\alpha) \,|\, N_0 = n-m\} = P\{Bi(n,\vartheta) > k_r(m,\alpha)\}$$
$$= P\Big\{\underbrace{\frac{Bi(m,\vartheta)-\mu}{\sigma}}_{\to N(0,1)} > \underbrace{\frac{k_r(m,\alpha)-\mu}{\sigma}}_{(*)}\Big\},$$

wobei $\mu = E(Bi(m,\vartheta)) = m\vartheta$ und $\sigma^2 = Var(Bi(m,\vartheta)) = m\vartheta(1-\vartheta)$. Es gilt

$$(*) \;=\; \frac{k_r(m,\alpha)-\mu}{\sigma} \;=\; \left(\frac{k_r-\mu_0}{\sigma_0} - \frac{\mu-\mu_0}{\sigma_0}\right)\frac{\sigma_0}{\sigma},$$

wobei $\mu_0 = E(Bi(m,\tfrac{1}{2}) = m/2$ und $\sigma_0^2 = Var(Bi(m,\tfrac{1}{2})) = m/4$. Aufgrund des zentralen Grenzwertsatzes gilt für $m \to \infty$

$$\frac{k_r-\mu_0}{\sigma_0} \;\to\; z_{1-\alpha},$$

und weiter ist

$$\vartheta = \frac{p_+}{p_+ + p_-} = \frac{p_+ + p_-}{2(p_+ + p_-)} + \frac{p_+ - p_-}{2(p_+ + p_-)} = \tfrac{1}{2} + \frac{d}{2(1-p_0)} = \tfrac{1}{2} + \frac{\delta}{2(1-p_0)\sqrt{n}}.$$

Somit erhalten wir bei unserem Grenzübergang

$$\vartheta \;\to\; \tfrac{1}{2},\; \frac{\sigma_0}{\sigma} = \frac{1}{2\sqrt{\vartheta(1-\vartheta)}} \;\to\; 1,$$

und es gilt

$$\frac{\mu-\mu_0}{\sigma_0} = \frac{m(\vartheta-\frac{1}{2})}{\sqrt{m/4}} = 2\sqrt{m}\,\frac{\delta}{2(1-p_0)\sqrt{n}} = \frac{\delta}{1-p_0}\sqrt{\frac{m}{n}}.$$

Es stellt sich die Frage, wohin der Quotient $m/n$ konvergiert bei unserem Grenzübergang. Da $N_0/n \xrightarrow{P} p_0$ aufgrund des schwachen Gesetzes der großen Zahlen, so ist es plausibel, dass wir uns bei der Bedingung $N_0 = n-m$ auf Werte von $m$ beschränken können, für welche $\lim(n-m)/n = p_0$, d.h. für welche $\lim m/n = 1-p_0$ gilt. Dann haben wir

$$\frac{\mu-\mu_0}{\sigma_0} \;=\; \frac{\delta}{1-p_0}\sqrt{\frac{m}{n}} \;\longrightarrow\; \frac{\delta}{\sqrt{1-p_0}},$$

und wir erhalten beim Grenzübergang $n \to \infty$ ($\delta$, $p_0$, $\alpha$ fest), $m/n \to 1-p_0$

$$(4)\quad \beta_{n|m}(\delta) \xrightarrow[n\to\infty]{} \beta(\delta) = P\left\{N(0,1) > z_{1-\alpha} - \frac{\delta}{\sqrt{1-p_0}}\right\} = P\left\{N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{1-p_0}}\right\}.$$

Nun wollen wir zeigen, dass die Grenzfunktion in (4) auch der Grenzwert der globalen (unbedingten) Gütefunktion (3) ist. Aufgrund des schwachen Gesetzes der großen Zahlen gilt $N_0/n \xrightarrow{P} p_0$, und somit existiert aufgrund des Satzes im Anhang A1 eine Folge $\varepsilon_n$ mit $\varepsilon_n \downarrow 0$, für welche gilt $P\{|N_0/n - p_0| > \varepsilon_n\} \to 0$ für $n \to \infty$. Es sei

$$M_n = \left\{m \,\middle|\, \left|\frac{n-m}{n} - p_0\right| \le \varepsilon_n\right\} = \left\{m \,\middle|\, \left|1 - p_0 - \frac{m}{n}\right| \le \varepsilon_n\right\} \text{ und } M_n^c = \left\{m \,\middle|\, \left|1 - p_0 - \frac{m}{n}\right| > \varepsilon_n\right\}.$$

Dann gilt für $n \to \infty$

$$P\{n - N_0 \in M_n\} \to 1 \quad \text{und} \quad P\{n - N_0 \in M_n^c\} \to 0.$$

Die globale (unbedingte) Gütefunktion (3) kann geschrieben werden als

$$\begin{aligned}\beta_n(\delta) &= P\{\text{Test B verwirft } H_0\} = \sum_{m=0}^{n} P\{N_0 = n-m\}\, \beta_{n|m}(\delta) \\ &= \sum_{m\in M_n} P\{N_0 = n-m\}\, \beta_{n|m}(\delta) + \sum_{m\in M_n^c} P\{N_0 = n-m\}\, \beta_{n|m}(\delta) = S_1 + S_2.\end{aligned} \tag{5}$$

Für die zweite Summe gilt

$$S_2 = \sum_{m\in M_n^c} P\{N_0 = n-m\}\, \beta_{n|m}(\delta) \le \sum_{m\in M_n^c} P\{N_0 = n-m\} \longrightarrow 0.$$

Nun betrachten wir die erste Summe $S_1$. Für $m \in M_n$ gilt

$$\lim_{n\to\infty} m/n = 1 - p_0,$$

und gemäß (4) haben wir dann

$$\lim_{n\to\infty} \beta_{n|m}(\delta) = \beta(\delta),$$

Somit gilt für die erste Summe

$$S_1 = \sum_{m\in M_n} P\{N_0 = n-m\}\, \beta_{n|m}(\delta) \longrightarrow \beta(\delta) \underbrace{\lim_{n\to\infty} P\{n - N_0 \in M_n\}}_{=1} = \beta(\delta).$$

Damit gilt auch für die globale Gütefunktion (3)

$$\beta_n(\delta) \xrightarrow[n\to\infty]{} \beta(\delta) = P\left\{N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{1-p_0}}\right\}. \tag{6}$$

## 4. Vergleich der asymptotischen Gütefunktionen

Jetzt wollen wir die asymptotischen Gütefunktionen zu den beiden Testverfahren miteinander vergleichen:

Test A: Beobachtungen mit $X_i = 0$ zufällig zuordnen nach + oder –;

Test B: Beobachtungen mit $X_i = 0$ unberücksichtigt lassen.

Wir haben oben unter (2) und (6) die folgenden asymptotischen Gütefunktionen gefunden :

Test A: $\beta_n^A(\delta) \xrightarrow[n\to\infty]{} \beta^A(\delta) = P\{N(0,1) < -z_{1-\alpha} + c_A\delta\}$ mit $c_A = 1$;

Test B: $\beta_n^B(\delta) \xrightarrow[n\to\infty]{} \beta^B(\delta) = P\{N(0,1) < -z_{1-\alpha} + c_B\delta\}$ mit $c_B = \dfrac{1}{\sqrt{1-p_0}}$.

Für $p_0 = 0$ ist $P\{N_0 > 0\} = 0$, und die beiden Tests sind dann identisch. Für $p_0 > 0$ ist der Koeffizient $c_B > 1$, und die Ablehnwahrscheinlichkeit bei Test B ist daher für $\delta > 0$ stets größer als bei Test A; damit ist Test B trennschärfer (effizienter) als Test A (für genügende große Stichprobenumfänge *n*). Die asymptotische relative Effizienz beträgt

$$are(B:A) = \lim \frac{n_A}{n_B} = \frac{c_B^2}{c_A^2} = \frac{1}{1-p_0} \quad (\geq 0)$$

(vgl. Kapitel 2). Für $p_0 = \frac{1}{2}$ z.B. ist $are(B:A) = 2$, und Test B ist doppelt so effizient wie Test A; mit $n_B = 100$ Beobachtungen erreicht man bei Test B praktisch die gleiche Trennschärfe wie mit $n_A = 200$ Beobachtungen bei Test A (Bestätigung mit Simulationsstudie).

## 5. Vergleich bei endlichem Stichprobenumfang

Wie wir oben gesehen haben, sind die exakten Gütefunktionen zu den beiden Tests A und B gegeben durch

$$\beta_n^A(\delta) = P\{\text{Test A verwirft } H_0\} = P\{Bi(n,p) > k_r(n,\alpha)\} \text{ mit } p = p_+ + \tfrac{1}{2}p_0$$

(7)

$$\beta_n^B(\delta) = P\{\text{Test B verwirft } H_0\} = \sum_{m=0}^{n} P\{N_0 = n-m\}\, P\{Bi(m,\vartheta) > k_r(m,\alpha)\} \text{ mit } \vartheta = \frac{p_+}{p_+ + p_-}$$

wobei $\delta = d\sqrt{n}$, $d = p_+ - p_-$, und es gilt für $n \to \infty$ ($\delta$, $p_0$, $\alpha$ fest)

$$\beta_n^A(\delta) \xrightarrow[n\to\infty]{} \beta^A(\delta) = P\{N(0,1) < -z_{1-\alpha} + c_A\delta\} \text{ mit } c_A = 1;$$

(8)

$$\beta_n^B(\delta) \xrightarrow[n\to\infty]{} \beta^B(\delta) = P\{N(0,1) < -z_{1-\alpha} + c_B\delta\} \text{ mit } c_B = \frac{1}{\sqrt{1-p_0}}.$$

Die Tabellen 1a, 1b und 1c zeigen die Werte der exakten Gütefunktionen sowie der asymptotischen Gütefunktionen für $n = 25, 100, 400$. In allen drei Tabellen sind die Werte von $\delta$ konstant gehalten (vgl Spalte 2); wenn der Stichprobenumfang *n* vervierfacht wird, halbieren sich die zugehörigen Werte des Parameters *d* (vgl. Spalte 1). Offensichtlich konvergieren die Werte der exakten Gütefunktion gegen die jeweiligen Werte der Grenzfunktion. Das gleiche Verhalten zeigt sich auch bei den Tabellen 2a, 2b und 2c, bei welchen die Stichprobenumfänge $n = 50, 200, 800$ betragen. Ferner kann man sehen, dass die Gütefunktion des Tests B als Funktion des Parameters *d* für $n = 25$ (bzw. $n = 100$ bzw. $n = 400$) nahezu übereinstimmt mit der Gütefunktion des Tests A für $n = 50$ (bzw. $n = 200$ bzw. $n = 800$). Dies ist eine Bestätigung der Tatsache, dass die asymptotische relative Effizienz des Test B im Vergleich zum Test A den Wert 2 besitzt für $p_0 = \frac{1}{2}$.

**Tabelle 1a:** Gütefunktionen für $n = 25,\ p_0 = 0.5,\ \alpha = 0.05,\ k_r = 17,\ \delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $d$ | $\delta$ | $p_-$ | $p_+$ | $p$ | $\vartheta$ | Test A | | Test B | |
| | | | | | | $\beta_n^A(\delta)$ | $\beta^A(\delta)$ | $\beta_n^B(\delta)$ | $\beta^B(\delta)$ |
| 0 | 0 | 0.25 | 0.25 | 0.5 | 0.5 | 0.0216 | 0.0500 | 0.0277 | 0.0500 |
| 0.1 | 0.5 | 0.2 | 0.3 | 0.55 | 0.6 | 0.0639 | 0.1261 | 0.1115 | 0.1742 |
| 0.2 | 1 | 0.15 | 0.35 | 0.6 | 0.7 | 0.1536 | 0.2595 | 0.3098 | 0.4088 |
| 0.3 | 1.5 | 0.1 | 0.4 | 0.65 | 0.8 | 0.3061 | 0.4424 | 0.6161 | 0.6831 |
| 0.4 | 2 | 0.05 | 0.45 | 0.7 | 0.9 | 0.5118 | 0.6388 | 0.8960 | 0.8817 |
| 0.5 | 2.5 | 0 | 0.5 | 0.75 | 1 | 0.7265 | 0.8038 | 0.9995 | 0.9707 |

**Tabelle 1b:** Gütefunktionen für $n = 100,\ p_0 = 0.5,\ \alpha = 0.05,\ k_r = 58,\ \delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $d$ | $\delta$ | $p_-$ | $p_+$ | $p$ | $\vartheta$ | Test A | | Test B | |
| | | | | | | $\beta_n^A(\delta)$ | $\beta^A(\delta)$ | $\beta_n^B(\delta)$ | $\beta^B(\delta)$ |
| 0 | 0 | 0.25 | 0.25 | 0.5 | 0.5 | 0.0443 | 0.0500 | 0.0377 | 0.0500 |
| 0.05 | 0.5 | 0.225 | 0.275 | 0.525 | 0.55 | 0.1145 | 0.1261 | 0.1416 | 0.1742 |
| 0.1 | 1 | 0.2 | 0.3 | 0.55 | 0.6 | 0.2415 | 0.2595 | 0.3589 | 0.4088 |
| 0.15 | 1.5 | 0.175 | 0.325 | 0.575 | 0.65 | 0.4218 | 0.4424 | 0.6429 | 0.6831 |
| 0.2 | 2 | 0.15 | 0.35 | 0.6 | 0.7 | 0.6225 | 0.6388 | 0.8686 | 0.8817 |
| 0.25 | 2.5 | 0.125 | 0.375 | 0.625 | 0.75 | 0.7964 | 0.8038 | 0.9722 | 0.9707 |
| 0.3 | 3 | 0.1 | 0.4 | 0.65 | 0.8 | 0.9123 | 0.9123 | 0.9972 | 0.9953 |
| 0.35 | 3.5 | 0.075 | 0.425 | 0.675 | 0.85 | 0.9709 | 0.9682 | 0.9999 | 0.9995 |
| 0.4 | 4 | 0.05 | 0.45 | 0.7 | 0.9 | 0.9928 | 0.9907 | 1.0000 | 1.0000 |

**Tabelle 1c:** Gütefunktionen für $n = 400,\ p_0 = 0.5,\ \alpha = 0.05,\ k_r = 216,\ \delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $d$ | $\delta$ | $p_-$ | $p_+$ | $p$ | $\vartheta$ | Test A | | Test B | |
| | | | | | | $\beta_n^A(\delta)$ | $\beta^A(\delta)$ | $\beta_n^B(\delta)$ | $\beta^B(\delta)$ |
| 0 | 0 | 0.25 | 0.25 | 0.5 | 0.5 | 0.0494 | 0.0500 | 0.0431 | 0.0500 |
| 0.025 | 0.5 | 0.2125 | 0.2625 | 0.5125 | 0.525 | 0.1250 | 0.1261 | 0.1566 | 0.1742 |
| 0.05 | 1 | 0.225 | 0.275 | 0.525 | 0.55 | 0.2578 | 0.2595 | 0.3819 | 0.4088 |
| 0.075 | 1.5 | 0.2125 | 0.2875 | 0.5375 | 0.575 | 0.4407 | 0.4424 | 0.6597 | 0.6831 |
| 0.1 | 2 | 0.2 | 0.3 | 0.55 | 0.6 | 0.6380 | 0.6388 | 0.8707 | 0.8817 |
| 0.125 | 2.5 | 0.1875 | 0.3125 | 0.5625 | 0.625 | 0.8043 | 0.8038 | 0.9683 | 0.9707 |
| 0.15 | 3 | 0.175 | 0.325 | 0.575 | 0.65 | 0.9136 | 0.9123 | 0.9952 | 0.9953 |
| 0.175 | 3.5 | 0.1625 | 0.3375 | 0.5875 | 0.675 | 0.9694 | 0.9682 | 0.9996 | 0.9995 |
| 0.2 | 4 | 0.15 | 0.35 | 0.6 | 0.7 | 0.9915 | 0.9907 | 1.0000 | 1.0000 |

**Tabelle 2a:** Gütefunktionen für $n = 50,\ p_0 = 0.5,\ \alpha = 0.05,\ k_r = 31,\ \delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $d$ | $\delta$ | $p_-$ | $p_+$ | $p$ | $\vartheta$ | Test A | | Test B | |
| | | | | | | $\beta_n^A(\delta)$ | $\beta^A(\delta)$ | $\beta_n^B(\delta)$ | $\beta^B(\delta)$ |
| 0 | 0 | 0.25 | 0.25 | 0.5 | 0.5 | 0.0325 | 0.0500 | 0.0336 | 0.0500 |
| 0.1 | 0.7071 | 0.2 | 0.3 | 0.55 | 0.6 | 0.1273 | 0.1742 | 0.2021 | 0.2595 |
| 0.2 | 1.4142 | 0.15 | 0.35 | 0.6 | 0.7 | 0.3356 | 0.4088 | 0.5813 | 0.6388 |
| 0.3 | 2.1213 | 0.1 | 0.4 | 0.65 | 0.8 | 0.6216 | 0.6831 | 0.9130 | 0.9123 |
| 0.4 | 2.8284 | 0.05 | 0.45 | 0.7 | 0.9 | 0.8594 | 0.8817 | 0.9975 | 0.9907 |
| 0.5 | 3.5355 | 0 | 0.5 | 0.75 | 1 | 0.9713 | 0.9707 | 1.0000 | 0.9996 |

**Tabelle 2b:** Gütefunktionen für $n = 200,\ p_0 = 0.5,\ \alpha = 0.05,\ \delta = d\sqrt{n},\ k_r = 112$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $d$ | $\delta$ | $p_-$ | $p_+$ | $p$ | $\vartheta$ | Test A | | Test B | |
| | | | | | | $\beta_n^A(\delta)$ | $\beta^A(\delta)$ | $\beta_n^B(\delta)$ | $\beta^B(\delta)$ |
| 0 | 0 | 0.25 | 0.25 | 0.5 | 0.5 | 0.0384 | 0.0500 | 0.0403 | 0.0500 |
| 0.05 | 0.7071 | 0.225 | 0.275 | 0.525 | 0.55 | 0.1441 | 0.1742 | 0.2273 | 0.2595 |
| 0.1 | 1.4142 | 0.2 | 0.3 | 0.55 | 0.6 | 0.3620 | 0.4088 | 0.6032 | 0.6388 |
| 0.15 | 2.1213 | 0.175 | 0.325 | 0.575 | 0.65 | 0.6408 | 0.6831 | 0.9022 | 0.9123 |
| 0.2 | 2.8284 | 0.15 | 0.35 | 0.6 | 0.7 | 0.8603 | 0.8817 | 0.9910 | 0.9907 |
| 0.25 | 3.5355 | 0.125 | 0.375 | 0.625 | 0.75 | 0.9651 | 0.9707 | 0.9998 | 0.9996 |
| 0.3 | 4.2426 | 0.1 | 0.4 | 0.65 | 0.8 | 0.9948 | 0.9953 | 1.0000 | 1.0000 |

**Tabelle 2c:** Gütefunktionen für $n = 800,\ p_0 = 0.5,\ \alpha = 0.05,\ \delta = d\sqrt{n},\ k_r = 423$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $d$ | $\delta$ | $p_-$ | $p_+$ | $p$ | $\vartheta$ | Test A | | Test B | |
| | | | | | | $\beta_n^A(\delta)$ | $\beta^A(\delta)$ | $\beta_n^B(\delta)$ | $\beta^B(\delta)$ |
| 0 | 0 | 0.25 | 0.25 | 0.5 | 0.5 | 0.0483 | 0.0500 | 0.0451 | 0.0500 |
| 0.025 | 0.7071 | 0.2375 | 0.2675 | 0.5125 | 0.525 | 0.1698 | 0.1742 | 0.2437 | 0.2595 |
| 0.05 | 1.4142 | 0.225 | 0.275 | 0.525 | 0.55 | 0.4024 | 0.4088 | 0.6210 | 0.6388 |
| 0.075 | 2.1213 | 0.2125 | 0.2875 | 0.5375 | 0.575 | 0.6778 | 0.6831 | 0.9059 | 0.9123 |
| 0.1 | 2.8284 | 0.2 | 0.3 | 0.55 | 0.6 | 0.8794 | 0.8817 | 0.9902 | 0.9907 |
| 0.125 | 3.5355 | 0.1875 | 0.3125 | 0.5625 | 0.625 | 0.9703 | 0.9707 | 0.9996 | 0.9996 |
| 0.15 | 4.2426 | 0.175 | 0.325 | 0.575 | 0.65 | 0.9954 | 0.9953 | 1.0000 | 1.0000 |

*Minitab-Makro* `minitab2.mac` *zur Berechnung der Gütefunktion* $\beta_n^B(\delta)$ *gemäß (7)*

```
gmacro
note
# Exakte Berechung der Gütefunktion beim Zeichentest
# (indifferente Fälle weglassen)
# H0: p_plus <= p_minus  gegen  H1: p_plus > p_minus
#
erase c1-c5 k5-k12
name c1 'm-Werte' c2 'kr(m,alpha)' c3 'prob_1' c4 'prob_2' c5 'product'
name k1 'n' k2 'alpha' k3 'p_minus' k4 'p_plus' k5 'p_null' k6 'p' k7 'theta'
name k8 'm' k9 'kr' k10 'summe' k11 'prob1' k12 'prob2' k13 'alphac'
note
note  Eingabewerte: K1 = n (Stichprobenumfang)
note                K2 = alpha = Signifikanzniveau
note                K3 = p_minus
note                K4 = p_plus
note
prin k1-k4
note Sind die Eingabewerte korrekt? (y/n)?
yesno k10
if k10 = 0
  exit
endif

let 'p_null'  = 1-'p_plus'-'p_minus'
let 'p' = 'p_plus'+'p_minus'
let 'theta' = 'p_plus'/'p'

let 'summe' = 0
do 'm'=1:'n'
  # für m=0 ist die Ablehnw'keit stets =0
  PDF 'm' 'prob1';
    Binomial 'n' 'p'.
  # kritische Schranke zu Bi(m,1/2)
  let 'alphac' = 1-'alpha'
  InvCDF 'alphac' 'kr';
    Binomial 'm' 0.5.
  CDF 'kr' 'prob2';
    Binomial 'm' 'theta'.
  let 'prob2' = 1-'prob2'
  let 'summe' = 'summe'+'prob1'*'prob2'
  let c1('m') = 'm'
  let c2('m') = 'kr'
  let c3('m') = 'prob1'
  let c4('m') = 'prob2'
  let c5('m') = 'prob1'*'prob2'
enddo
prin summe
note summe = Ablehnwahrscheinlichkeit (Wert der Gütefunktion)
endmacro
```

# 5. Test von McNemar

## 1. Problemstellung, Variante 1

Lit.: Pfanzagl, J. (1972). Allgemeine Methodenlehre der Statistik II. Sammlung Göschen - de Gruyter, Berlin. Seite 139 ff.

Wir möchten wissen, welche der beiden Abituraufgaben A und B schwieriger zu lösen ist. Dazu wird folgendes Experiment durchgeführt. Es werden $n = 100$ Abiturienten (Versuchspersonen) zufällig ausgewählt. Jeder Versuchperson werden die beiden Aufgaben A und B zur Lösung vorgelegt. Das Ergebnis ist gegeben in Tabelle 1; von den $n = 100$ Abiturienten haben $X_A = 74$ die Aufgabe A gelöst, jedoch $X_B = 83$ die Aufgabe B. Es scheint also, dass Aufgabe A schwieriger ist als Aufgabe B. Ist das Ergebnis unseres Experiments ein statistischer Beweis, dass Aufgabe A schwieriger ist als Aufgabe B, oder bewegt sich die beobachtete Differenz der Häufigkeiten im Rahmen des Zufälligen? Um diese Frage zu beantworten, benötigen wird ein adäquates statistisches Modell zu unserem Experiment. Zunächst betrachten wir ein Modell, das nicht korrekt ist, und das zu einem naheliegenden Trugschluss führt, und dann untersuchen wir das korrekte Modell.

**Tabelle 1:** Ergebnis zum Abiturexperiment (einfache Aufgliederung)

| | + | – | $\sum$ |
|---|---|---|---|
| $X_A$ | 74 | 26 | 100 |
| $X_B$ | 83 | 17 | 100 |
| $\sum$ | 157 | 43 | 200 |

*Naheliegendes Modell (nicht korrekt):*

$X_A$ , $X_B$ unabhängige Zufallsgrößen je mit einer Binomialverteilung $Bi(n, p_A)$ bzw. $Bi(n, p_B)$;

(1) $H_0$: $p_A = p_B$ , d.h. Aufgabe A und B sind gleich schwierig;

$H_1$: $p_A < p_B$ , d.h. Aufgabe A ist schwieriger als Aufgabe B.

Der korrekte Test zu diesem Testproblem ist der Vierfelder-Test von Fisher; der exakte p-Wert dieses Tests ist gegeben durch

$$p_{crit} = P\{H(200,157,100) \le 74\} = 0.0840,$$

wobei die Schreibweise $P\{H(N,M,n) \ge k\}$ äquivalent ist zu

$$P\{Y > k\} \text{ mit } Y \sim H(N,M,n);$$

dabei bezeichnet wir mit $H(N,M,n)$ die hypergeometrischen Verteilung mit den Parametern $N = 200$ (= Anzahl der Kugeln in der Urne), $M = 157$ (Anzahl der weißen Kugeln in der Urne) und mit $n = 100$ (= Anzahl der ausgewählten Kugeln ohne Zurücklegen), und die Zufallsgröße $Y$ ist die Anzahl der weißen Kugeln in der Stichprobe. Bei einem vorgegebenen Signifikanzniveau von $\alpha = 0.05$ würde also das vorliegende Ergebnis nicht zur Ablehnung der Nullhypothese führen. Das vorliegende *Modell ist jedoch nicht korrekt*, denn die beiden Zufallsgrößen $X_A$ und $X_B$ sind offensichtlich *nicht stochastisch unabhängig*. Die beiden Aufgaben sind ja nicht zwei unabhängigen Stichproben mit je $n = 100$ Versuchspersonen vorgelegt worden, sondern den gleichen $n = 100$ Versuchsperonen; jemand der Aufgabe A löst, wird Aufgabe B eher lösen, als jemand der Aufgabe A nicht löst, da er ja im ersten Fall vermutlich intelligenter ist als im zweiten Fall.

*Korrektes Modell:*

Die Population der Abiturienten besteht aus vier Kategorien:

Kategorie $(+,+)=$ Abiturienten, welche beide Aufgaben A und B lösen;

Kategorie $(+,-)=$ Abiturienten, welche Aufgabe A lösen und Aufgabe B nicht lösen;

Kategorie $(-,+)=$ Abiturienten, welche Aufgabe A nicht lösen und Aufgabe B lösen;

Kategorie $(-,-)=$ Abiturienten, welche beide Aufgaben A und B nicht lösen.

Wir bezeichnen mit

$p_{11}=$ Anteil der Abiturienten in Kategorie $(+,+)$;

$p_{12}=$ Anteil der Abiturienten in Kategorie $(+,-)$;

$p_{21}=$ Anteil der Abiturienten in Kategorie $(-,+)$;

$p_{22}=$ Anteil der Abiturienten in Kategorie $(-,-)$.

$p_{11}$ ist also die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter Abiturient zur Kategorie $(+,+)$ gehört (analog für die übrigen $p_{ij}$). Die Population der Abiturienten kann also beschrieben werden durch die Wahrscheinlichkeiten in Tabelle 2.

**Tabelle 2:** Population der Abiturienten (zweidimensionale Aufgliederung)

| A \ B | + | − | |
|---|---|---|---|
| + | $p_{11}$ | $p_{12}$ | $p_{1\bullet}$ |
| − | $p_{21}$ | $p_{22}$ | $p_{2\bullet}$ |
| | $p_{\bullet 1}$ | $p_{\bullet 2}$ | 1 |

**Tabelle 3:** Stichprobe vom Umfang $n$ (zweidimensionale Aufgliederung)

| A \ B | + | − | |
|---|---|---|---|
| + | $X_{11}$ | $X_{12}$ | $X_{1\bullet}$ |
| − | $X_{21}$ | $X_{22}$ | $X_{2\bullet}$ |
| | $X_{\bullet 1}$ | $X_{\bullet 2}$ | $n$ |

Nun wählen wir aus unserer Population $n=100$ Abiturienten zufällig aus, und bestimmen die folgenden Häufigkeiten:

$X_{11}=$ Anzahl der Abiturienten der Kategorie $(+,+)$, d.h. Abiturienten, welche A und B lösen;

$X_{12}=$ Anzahl der Abiturienten der Kategorie $(+,-)$;

$X_{21}=$ Anzahl der Abiturienten der Kategorie $(-,+)$;

$X_{22}=$ Anzahl der Abiturienten der Kategorie $(-,-)$.

Wenn wir annehmen, dass die Anzahl der Abiturienten eines bestimmten Jahrganges groß ist (statische Interpretation), bzw. dass das Ausbildungsniveau unverändert bleibt, sodass sich die Auswahlwahrscheinlichkeiten $p_{ij}$ nicht ändern (dynamische Interpretation), so gilt

(2) $X=(X_{11},X_{12},X_{21},X_{22}) \sim \mathrm{Mu}(n,p_{11},p_{12},p_{21},p_{22})$,

wobei wir mit $Mu(n,\eta_1,\ldots,\eta_k)$ die Multinomialverteilung bezeichnen (Urnenmodell mit Zurücklegen, $k$ Kategorien mit den Auswahlwahrscheinlichkeiten $\eta_1,\ldots,\eta_k$). Unser Testproblem hat jetzt die Form

(3) $\mathrm{H}_0: p_{12}=p_{21}$, d.h. Aufgabe A und B sind gleich schwierig;
$\mathrm{H}_1: p_{12}<p_{21}$, d.h. Aufgabe A ist schwieriger als Aufgabe B.

Die Nullhypothese besagt, dass die Kontingenztafel $(p_{ij})$ *symmetrisch* ist, und daher spricht man beim vorliegenden Testproblem auch von der *Symmetrieprüfung in einer Kontingenztafel*. Es sei

$p_A=p_{1\bullet}=p_{11}+p_{12}=$ W'keit, dass eine zufällig ausgewählte Vp. Aufgabe A löst;

$p_B=p_{\bullet 1}=p_{11}+p_{21}=$ W'keit, dass eine zufällig ausgewählte Vp. Aufgabe B löst.

Das Testproblem (3) kann dann auch geschrieben werden als

$H_0 : p_A = p_B$, d.h. Aufgabe A und B sind gleich schwierig;

$H_1 : p_A < p_B$, d.h. Aufgabe A ist schwieriger als Aufgabe B;

dies ist die gleiche Form wie bei (1). Nun gilt

$$X_{12} \big| X_{12} + X_{21} = m \;\sim\; Bi\big(m, \vartheta\big), \text{ wobei } \vartheta = \frac{p_{12}}{p_{12} + p_{21}};$$

(4)

$H_0 : p_{12} = p_{21} \;\Leftrightarrow\; \vartheta = \frac{1}{2}$;

$H_1 : p_{12} < p_{21} \;\Leftrightarrow\; \vartheta < \frac{1}{2}$.

Die bedingte Verteilung von $X_{12}$ bei gegebener Summe $X_{12} + X_{21} = m$ ist also eine Binomialverteilung $Bi(m, \vartheta)$, und das Testproblem (3) ist jetzt äquivalent zu

$H_0 : \vartheta = \frac{1}{2}$, d.h. Aufgabe A und B sind gleich schwierig;

$H_1 : \vartheta < \frac{1}{2}$, d.h. Aufgabe A ist schwieriger als Aufgabe B.

Der Test von McNemar besitzt nun die folgende Entscheidungsregel:

Berechne $m = X_{12} + X_{21}$;

bestimme $k_\ell =$ linke kritische $\alpha$-Schranke zu $Bi(m, \frac{1}{2})$;

verwirf $H_0$, falls $X_{12} < k_\ell$.

Diese Entscheidungsregel kann in äquivalenter Form auch mit Hilfe des sogenannten p-Werts formuliert werden:

Es seien $k_{12}$ und $k_{21}$ die Realisationen zu $X_{12}$ und $X_{21}$ und $m = k_{12} + k_{21}$;

berechne $p_{\text{crit}} = P\left\{ Bi(m, \frac{1}{2}) \le k_{12} \right\}$;

verwirf $H_0$, falls $p_{\text{crit}} \le \alpha$.

Beim Test von McNemar sind nur die Werte $k_{12}$ und $k_{21}$ der Nebendiagonalen relevant für die Testentscheidung. Wir werden zeigen, dass der Test von McNemar auch als globaler (unbedingter) Test das Signifikanzniveau $\alpha$ stets einhält, und dass der Test konsistent ist. Bei der einseitigen Fragestellung in die andere Richtung sowie bei der zweiseitigen Fragestellung gilt alles in analoger Weise.

*Beispiel*

Wir kehren zurück zum Beispiel der Tabelle 1 und interessieren uns für die zweidimensionale Aufgliederung der Daten; die Ergebnisse der Tabelle 1 führen zu folgender Randverteilung bei der zweidimensionalen Aufgliederung:

**Tabelle 4:** Randverteilung beim Abiturexperiment
(zweidimensionale Aufgliederung)

| A \ B | + | – | |
|---|---|---|---|
| + | x | x | 74 |
| – | x | x | 26 |
| | 83 | 17 | 100 |

Wir benötigen nun zusätzlich die Information über die Besetzung der vier Felder der Tabelle 4. Dafür gibt es zwei Grenzfälle:

**Tabelle 4a:** Grenzfall 1

| A \ B | + | – | |
|---|---|---|---|
| + | 57 | 17 | 74 |
| – | 26 | 0 | 26 |
| | 83 | 17 | 100 |

**Tabelle 4b:** Grenzfall 2

| A \ B | + | – | |
|---|---|---|---|
| + | 74 | 0 | 74 |
| – | 9 | 17 | 26 |
| | 83 | 17 | 100 |

Für die beiden Grenzfälle finden wir die folgenden p-Werte:

Grenzfall 1: $m = k_{12} + k_{21} = 17 + 26 = 43$, $p_{\text{crit}} = P\left\{Bi(43, \frac{1}{2}) \le 17\right\} = 0.1110$;

Grenzfall 2: $m = k_{12} + k_{21} = 0 + 9 = 9$, $p_{\text{crit}} = P\left\{Bi(9, \frac{1}{2}) \le 0\right\} = 0.00195$.

Es ist bemerkenswert, dass bei gegebener Randverteilung der p-Wert umso kleiner wird je kleiner die Besetzungszahlen der Nebendiagnonalen sind. Bei einem vorgegebenen Signifikanzniveau von $\alpha = 0.05$ kann die Nullhypothese im ersten Grenzfall nicht abgelehnt werden, wohl aber im zweiten.

## 2. Problemstellung, Variante 2

Wir möchten wissen, welche Veränderung eine Wahlkampagne bei den Wahlberechtigten bewirkt, oder welche Veränderung eine medizinische Behandlung bei bestimmten Versuchspersonen bewirkt. Wir wählen hier die Einkleidung des Problems in die Untersuchung der Auswirkung einer Wahlkampagne. Es wird folgendes Experiment durchgeführt. Es werden $n = 100$ Wahlberechtigte (Versuchspersonen) zufällig ausgewählt und sowohl vor der Wahlkampagne (A = vorher) als auch nach der Wahlkampagne (B = nacher) nach ihrer Einstellung zur Partei y befragt (+ = dafür, positiv; – = dagegen, negativ). Es sei

$p_{11} =$ Wahrscheinlichkeit, daß eine zufällig ausgewählte Vp. zur Kategorie (+,+) gehört, d.h. vorher und nachher "positiv" ist;

$p_{12} =$ Wahrscheinlichkeit, daß eine zufällig ausgewählte Vp. zur Kategorie (+,–) gehört, d.h. vorher "positiv" und nachher "negativ" ist;

etc.

und weiter

$X_{11} =$ Anzahl der zufällig ausgewählten Vp., welche zur Kategorie (+,+) gehören;

$X_{12} =$ Anzahl der zufällig ausgewählten Vp., welche zur Kategorie (+,–) gehören;

etc.

Die Wahrscheinlichkeiten $p_{ij}$ und die Häufigkeiten $X_{ij}$ können also wie in Tabelle 2 und 3 dargestellt werden. Falls die Gesamtzahl der Wahlberechtigten groß ist im Vergleich zum Stichprobenumfang *n*, so gilt

$$X = (X_{11}, X_{12}, X_{21}, X_{22}) \sim Mu(n, p_{11}, p_{12}, p_{21}, p_{22}).$$

Die Wahlkampagne ist erfolgreich, wenn nach der Kampagne mehr Wähler für die Partei y sind als vorher; nun ist

$p_A = p_{11} + p_{12} = p_{1\bullet} =$ W'keit, dass eine zufällig ausgewählte Vp. *vorher* für die Partei y ist

$=$ Anteil der Wahlberechtigten, welche *vorher* für die Partei y sind;

$p_B = p_{11} + p_{21} = p_{\bullet 1} =$ W'keit, dass eine zufällig ausgewählte Vp. *nachher* für die Partei y ist

$=$ Anteil der Wahlberechtigten, welche *nachher* für die Partei y sind.

Wir interessieren uns also hier für das Testproblem

$H_0$: $p_A = p_B$, d.h. die Wahlkampagne zeigt keine Wirkung,

$H_1$: $p_A < p_B$, d.h. die Wahlkampagne zeigt eine positive Wirkung,

und dieses Problem ist offensichtlich äquivalent zu

$H_0$: $p_{12} = p_{21}$ gegen $H_0$: $p_{12} < p_{21}$.

Wir haben also hier genau das gleiche formale Problem wie bei der ersten Variante unseres Problems, und somit kommt auch hier der Test von McNemar als Lösung in Frage.

## 3. Eigenschaften des Tests von McNemar

Beim Test von McNemar wird der kritische Wert $k_\ell$ bzw. die kritische Wahrscheinlichkeit $p_{\text{crit}}$ (vgl. Abschnitt 1) mit Hilfe der bedingten Verteilung von $X_{12}$ gegeben $X_{12}+X_{21}=m$ bestimmt (unter der Nullhypothese). Es stellt sich die Frage, ob ein vorgegebenes Signifikanzniveau $\alpha$ auch beim globalen (unbedingten) Test eingehalten wird. Aufgrund von (4) gilt

$$X_{12}\big|X_{12}+X_{21}=m \;\sim\; Bi\big(m,\vartheta\big),\;\text{ wobei }\;\vartheta=\frac{p_{12}}{p_{12}+p_{21}},$$

und somit kann die Wahrscheinlichkeit, dass der Test von McNemar die Nullhypothese ablehnt, geschrieben werden als

$$\begin{aligned}P\big\{&\text{Test von McNemar verwirft } \mathrm{H}_0\big\}=\\ &=\sum_{m=0}^{n} P\big\{X_{12}+X_{21}=m\big\}\,P\big\{\text{Test von McNemar verwirft } \mathrm{H}_0\,\big|\,X_{12}+X_{21}=m\big\}\\ &=\sum_{m=0}^{n} P\big\{X_{12}+X_{21}=m\big\}\,P\big\{Bi(m,\vartheta)<k_\ell\big\}.\end{aligned}$$

Nun gilt $X_+ = X_{12}+X_{21}\sim Bi(n,p_+)$ mit $p_+ = p_{12}+p_{21}$, und somit ist die Gütefunktion des Test von McNemar abhängig von den beiden Parametern $\vartheta$ und $p_+$, und sie kann geschrieben werden als

$$\begin{aligned}\beta_n(\vartheta,p_+) &= P\big\{\text{Test von McNemar verwirft } \mathrm{H}_0\big\}\\ &=\sum_{m=0}^{n} P\big\{Bi(n,p_+)=m\big\}\,P\big\{Bi(m,\vartheta)<k_\ell\big\}.\end{aligned}\tag{5}$$

Unter der Nullhypothese gilt $p_{12}=p_{21}$ d.h. $\vartheta=\frac{1}{2}$, und es gilt dann

$$\beta_n\Big(\tfrac{1}{2},p_+\Big) \;=\; \sum_{m=0}^{n} P\big\{Bi(n,p_+)=m\big\}\underbrace{P\Big\{Bi(m,\tfrac{1}{2})<k_\ell\Big\}}_{\leq\alpha \text{ für jedes } m} \;\leq\; \alpha\sum_{m=0}^{n} P\big\{Bi(n,p_+)=m\big\} \;=\; \alpha.$$

Die Irrtumswahrscheinlichkeit 1. Art ist also auch beim globalen (unbedingten) Test stets $\leq\alpha$.

Wir wollen weiter zeigen, dass der Test von McNemar ein konsistenter Test ist, dass also die Gütefunktion in jedem Punkt der Alternative für wachsenden Stichprobenumfang *n* gegen eins konvergiert. Da $X_+ = X_{12}+X_{21}\sim Bi(n,p_+)$ mit $p_+ = p_{12}+p_{21}$, so haben wir $X_+/n \xrightarrow{\;P\;} p_+$. Aufgrund des Satzes in Anhang A1 existiert eine Folge $\varepsilon_n$ mit $\varepsilon_n\downarrow 0$ und mit $P\Big\{\big|X_+/n-p_+\big|>\varepsilon_n\Big\}\to 0$ für $n\to\infty$. Nun sei

$$M_n=\left\{m\,\middle|\,\left|\frac{m}{n}-p_+\right|\leq\varepsilon_n\right\}\quad\text{und}\quad M_n^c=\left\{m\,\middle|\,\left|\frac{m}{n}-p_+\right|>\varepsilon_n\right\}.$$

Dann gilt für $n\to\infty$

$$P\big\{X_+\in M_n\big\}\to 1\quad\text{und}\quad P\Big\{X_+\in M_n^c\Big\}\to 0.\tag{6}$$

Die Gütefunktion (5) kann geschrieben werden als

$$\begin{aligned}\beta_n(\vartheta,p_+)&=\sum_{m=0}^{n} P\big\{Bi(n,p_+)=m\big\}\,P\big\{Bi(m,\vartheta)<k_\ell\big\}\\ &=\sum_{m\in M_n} P\big\{Bi(n,p_+)=m\big\}\,P\big\{Bi(m,\vartheta)<k_\ell\big\}+\sum_{m\in M_n^c}\cdots \;=\; S_1+S_2.\end{aligned}$$

Für die zweite Summe gilt

$$S_2 = \sum_{m \in M_n^c} P\{Bi(n, p_+) = m\}\, P\{Bi(m, \vartheta) < k_\ell\}$$

$$\leq \sum_{m \in M_n^c} P\{Bi(n, p_+) = m\} = P\left\{X_+ \in M_n^c\right\} \longrightarrow 0 \text{ für } n \to \infty.$$

Für die erste Summe ist $m \in M_n$ und offensichtlich gilt dann $m \to \infty$ für $n \to \infty$. Nun ist der gewöhnliche Binomialtest konsistent, und somit gilt

$$P\{Bi(m, \vartheta) < k_\ell\} \xrightarrow[m \to \infty]{} 1 \text{ für jedes } \vartheta < \tfrac{1}{2};$$

für jedes $\vartheta < \frac{1}{2}$ folgt daraus zusammen mit (6), dass $S_1 \to 1$ für $n \to \infty$ und weiter $\beta_n(\vartheta, p_+) \to 1$, d.h. der Test von McNemar ist konsistent.

# 6. Unabhängige oder verbundene Stichproben? Vier-Felder-Test oder Test von McNemar?

## 1. Problemstellung

Eine Population wird einer Behandlung unterzogen (A=vorher, B=nachher):

| A \ B | + | − | |
|---|---|---|---|
| + | $p_{11}$ | $p_{12}$ | $p_{1\bullet}$ |
| − | $p_{21}$ | $p_{22}$ | $p_{2\bullet}$ |
| | $p_{\bullet 1}$ | $p_{\bullet 2}$ | 1 |

$p_{11} =$ Wahrscheinlichkeit, daß eine (zufällig ausgewählte) Vp. vorher und nachher positiv ist;
$p_{12} =$ Wahrscheinlichkeit, daß eine (zufällig ausgewählte) Vp. vorher positiv und nachher negativ ist;
$p_{21} = \cdots$;
$p_{22} = \cdots$;

Wir interessieren uns für das Testproblem

(1) $H_0: p_{12} = p_{21}$, d.h. Behandlung hat keine Wirkung;
$H_1: p_{12} > p_{21}$, d.h. nach der Behandlung weniger Versuchspersonen positiv als vorher.

Dieses Testproblem können wir auf zwei unterschiedliche Arten angehen:

A) Unabhängige Stichproben, Vier-Felder-Test (vgl. Abschnitt 2)

B) Verbundene Stichproben, Test von McNemar (vgl. Abschnitt 3)

Um die beiden Verfahren zu vergleichen, berechnen wir jeweils die asymptotische Gütefunktion des betreffenden Tests, und im Abschnitt 4 vergleichen wir die asymptotischen Gütefunktionen

Wir betrachten im Folgenden nur das einseitige Testproblem (1); bei der einseitigen Fragestellung in die andere Richtung und bei der zweiseitigen Fragestellung ist alles analog.

## 2. Versuchsplan mit unabhängigen Stichproben; Vier-Felder-Test

Vorher und nachher je *n* Versuchspersonen zufällig auswählen und befragen (zwei unabhängige Stichproben je mit Umfang *n*);

$X_1 =$ Anzahl der Versuchspersonen, welche vorher positiv sind ;
$X_2 =$ Anzahl der Versuchspersonen, welche nachher positiv sind .

*Modell:*
$X_1 \sim Bi(n, \vartheta_1)$, $\vartheta_1 = p_{1\bullet} = p_{11} + p_{12}$;
$X_2 \sim Bi(n, \vartheta_2)$, $\vartheta_2 = p_{\bullet 1} = p_{11} + p_{21}$;
$X_1, X_2$ unabhängig.
$H_0: p_{12} \le p_{21} \quad \Leftrightarrow \quad \vartheta_1 \le \vartheta_2$
$H_1: p_{12} > p_{21} \quad \Leftrightarrow \quad \vartheta_1 > \vartheta_2$ .

Vierfeldertabelle:

| | + | − | |
|---|---|---|---|
| $X_1$ | $k_1$ | · | $n$ |
| $X_2$ | $k_2$ | · | $n$ |
| $X$ | $k$ | · | $2n$ |

*Exakter Vier-Felder-Test von Fisher:*

Falls $\vartheta_1 = \vartheta_2$ d.h. falls $p_{12} = p_{21}$, so gilt $X_1 \big| X = k \sim H(2n, n, k)$.

Es sei $k_r =$ rechte kritische $\alpha$-Schranke zu $H(2n, n, k)$.

Verwirf $H_0$, falls $X_1 > k_r$ .

*Gütefunktion:*

$$(2)\quad \beta_n(\vartheta_1,\vartheta_2) = P\{\text{Test verwirft } H_0\} = \sum_{k=0}^{2n} P\{X=k\}\,P\{\text{Test verwift } H_0 \mid X=k\}.$$

Nun ist der Vier-Felder-Test konsistent, d.h. für $\vartheta_1 > \vartheta_2$ gilt $\beta_n(\vartheta_1,\vartheta_2)\to 1$. Um eine nicht-ausgeartete Grenzfunktion zu erhalten, führen wir anstelle der ursprünglichen Variablen $\vartheta_1, \vartheta_2$ die neuen Variablen $\delta$ und $\bar{\vartheta}$ ein:

$$\delta = \sqrt{n}(\vartheta_1-\vartheta_2) = \sqrt{n}(p_{12}-p_{21})$$
$$\bar{\vartheta} = \tfrac{1}{2}(\vartheta_1+\vartheta_2) = p_{11}+\tfrac{1}{2}(p_{12}+p_{21}).$$

Die Gütefunktion ist dann abhängig von $n, \delta, \bar{\vartheta}, \alpha$. Für $n\to\infty$ $\left(\delta, \bar{\vartheta}, \alpha \text{ fest}\right)$ gilt (vgl. Anhang A2):

$$(3)\quad \beta_n(\delta) \longrightarrow \beta(\delta) = P\left\{N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{2\bar{\vartheta}(1-\bar{\vartheta})}}\right\}.$$

## 3. Versuchsplan mit verbundenen Stichproben; Test von McNemar

*n* Versuchspersonen zufällig auswählen und jede Vp. zweimal befragen (zwei verbundene Stichproben je mit Umfang *n*; A = vorher, B = nachher);

Stichprobe:

| A \ B | + | − | |
|---|---|---|---|
| + | $X_{11}$ | $X_{12}$ | |
| − | $X_{21}$ | $X_{22}$ | |
| | | | $n$ |

$X_{11}=$ Anzahl der Vp., welche vorher und nacher positiv sind;
$X_{12}=$ Anzahl der Vp., welche vorher positiv und nacher negativ sind;
$X_{21}=\cdots$;
$X_{22}=\cdots$.

*Modell:*
$X=(X_{11},X_{12},X_{21},X_{22}) \sim Mu(n, p_{11}, p_{12}, p_{21}, p_{22})$;
$H_0$: $p_{12}=p_{21}$ gegen $H_1$: $p_{12}>p_{21}$.

*Test von McNemar:*
Es gilt $X_+ = X_{12}+X_{21} \sim Bi(n, p_+)$ mit $p_+ = p_{12}+p_{21}$ und

$$X_{12}\,|\,X_+=m \sim Bi(m,\eta) \quad \text{mit} \quad \eta=\frac{p_{12}}{p_+}.$$

Falls $p_{12}=p_{21}$, so gilt:

$$X_{12}\,|\,X_+=m \sim Bi\left(m,\tfrac{1}{2}\right).$$

Es sei $k_r = k_r(m,\alpha)$ die rechte kritische $\alpha$-Schranke zu $Bi\left(m,\frac{1}{2}\right)$.

Verwirf $H_0$, falls $X_{12} > k_r$.

*Gütefunktion:*

$$\beta_n(p_{12},p_{21}) = P\{\text{Test verwirft } H_0\} = \sum_{m=0}^{n} P\{X_+=m\}\,P\{\text{Test verwift } H_0 \mid X_+=m\}$$

(4)

$$= \sum_{m=0}^{n} P\{X_+=m\}\,P\{Bi(m,\eta)>k_r\} \quad \text{mit} \quad \eta=\frac{p_{12}}{p_+}.$$

Nun ist der Test von McNemar konsistent, d.h. für $p_{12}>p_{21}$ gilt $\beta_n(p_{12},p_{21})\to 1$. Um eine nicht-ausgeartete Grenzfunktion zu erhalten, führen wir anstelle der ursprünglichen Variablen $p_{12}, p_{21}$

die neuen Variablen $\delta$ und $p_+$ ein:

$\delta = \sqrt{n}(p_{12} - p_{21})$ (wie beim Vier-Felder-Test)

$p_+ = p_{12} + p_{21}$.

Es gilt dann

$$p_{12} = \tfrac{1}{2} p_+ + \frac{\delta}{2\sqrt{n}} \quad \text{und} \quad p_{21} = \tfrac{1}{2} p_+ - \frac{\delta}{2\sqrt{n}}.$$

Die Gütefunktion ist abhängig von $(n, \delta, p_+, \alpha)$, und wir schreiben jetzt

$$\beta_n(\delta) = \mathrm{P}\{\text{Test von McNemar verwirft } \mathrm{H}_0\}$$

$$(5) \qquad = \sum_{m=0}^{n} P\{X_+ = m\} P\{\text{Test verwirft } \mathrm{H}_0 \mid X_+ = m\}$$

$$= \sum_{m=0}^{n} P\{X_+ = m\} \underbrace{P\{Bi(m,\eta) > k_r\}}_{= \beta_{n|m}(\delta)}.$$

Nun ist $X_+ \sim Bi(n, p_+)$, und daher gilt $X_+/n \xrightarrow{P} p_+$ für $n \to \infty$. Mit Hilfe des Satzes in Anhang A1 kann man zeigen (analog wie beim Vorzeichentest, vgl. Kapitel 4, Abschnitt 3), dass der

Grenzwert von $\beta_n(\delta)$ für $n \to \infty$ $(\delta, p_+, \alpha$ fest)

(6) gleich ist dem

Grenzwert von $\beta_{n|m}(\delta)$ für $n \to \infty, m/n \to p_+$ $(\delta, p_+, \alpha$ fest).

Wir betrachten daher bei der bedingten Gütefunktion $\beta_{n|m}(\delta)$ den Grenzübergang $n \to \infty$, $m/n \to p_+$ $(\delta, p_+, \alpha$ fest). Wir schreiben

$$(7) \quad \beta_{n|m}(\delta) = P\{Bi(m,\eta) > k_r\} = P\left\{\frac{Bi(m,\eta) - \mu}{\sigma} > \frac{k_r - \mu}{\sigma}\right\} = P\{(a) > (b)\},$$

wobei $\mu = E(Bi(m,\eta)) = m\eta$ und $\sigma^2 = Var(Bi(m,\eta)) = m\eta(1-\eta)$. Es gilt

$$\eta = \frac{p_{12}}{p_+} = \tfrac{1}{2} + \frac{\delta}{2\,p_+\sqrt{n}} \to \tfrac{1}{2} \quad \text{für } n \to \infty.$$

Für die standardisierte Binomialverteilung gilt

$$\frac{Bi(m,\eta) - \mu}{\sigma} \xrightarrow{d} N(0,1).$$

Nun ist

$$(8) \quad (b) = \frac{k_r - \mu}{\sigma} = \left(\frac{k_r - \mu_0}{\sigma_0} - \frac{\mu - \mu_0}{\sigma_0}\right)\frac{\sigma_0}{\sigma},$$

wobei $\mu_0 = E(Bi(m,\tfrac{1}{2})) = m/2$ und $\sigma_0^2 = Var(Bi(m,\tfrac{1}{2})) = m/4$. Da $\sigma_0/\sigma \to 1$, $(k_r - \mu_0)/\sigma_0 \to z_{1-\alpha}$ und

$$\frac{\mu - \mu_0}{\sigma_0} = 2\sqrt{m}(\eta - \tfrac{1}{2}) = 2\sqrt{m}\frac{\delta}{2\,p_+\sqrt{n}} = \sqrt{m/n}\,\frac{\delta}{p_+} \longrightarrow \sqrt{p_+}\,\frac{\delta}{p_+} = \frac{\delta}{\sqrt{p_+}},$$

so erhalten wir

$$(b) = \frac{k_r - \mu}{\sigma} \longrightarrow z_{1-\alpha} - \frac{\delta}{\sqrt{p_+}}.$$

Daher gilt aufgrund von (7)

$$\beta_{n|m}(\delta) \longrightarrow P\left\{N(0,1) > z_{1-\alpha} - \frac{\delta}{\sqrt{p_+}}\right\} = P\left\{N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{p_+}}\right\},$$

und weiter aufgrund von (6)

(9) $\beta_n(\delta) \longrightarrow \beta(\delta) = P\left\{N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{p_+}}\right\}$ für $n \to \infty$ $(\delta,\ p_+, \alpha$ fest$)$.

## 4. Asymptotische relative Effizienz

Wie im Kapitel 2 beim Vergleich des t-Tests mit dem Vorzeichentest wollen wir jetzt aufgrund der asymptotischen Gütefunktionen die asymptotische relative Effizienz bestimmen.

Population:

| A \ B | + | – |
|---|---|---|
| + | $p_{11}$ | $p_{12}$ |
| – | $p_{21}$ | $p_{22}$ |

$\delta = \sqrt{n}(p_{12} - p_{21})$

$H_0 : p_{12} = p_{21}$ d.h. $\delta = 0$

$H_1 : p_{12} > p_{21}$ d.h. $\delta > 0$

1. Versuchsplan: Unabhängige Stichproben mit Vier-Felder-Test (U)

2. Versuchplan: Verbundene Stichproben mit Test von McNemar (V)

*Asymptotische Gütefunktionen:*
Wir betrachten den Grenzübergang $n \to \infty$ $(p_{11},\ p_{22}, \delta$ und $\alpha$ fest$)$.

Dann gilt:

$p_+ = p_{12} + p_{21} = 1 - p_{11} - p_{22}$ bleibt fest beim Grenzübergang;

$\bar{\vartheta} = p_{11} + \frac{1}{2}(p_{12} + p_{21}) = p_{11} + \frac{1}{2}p_+$ bleibt fest beim Grenzübergang;

$$\beta_n^U(\delta) \longrightarrow \beta^U(\delta) = P\left\{N(0,1) < -z_{1-\alpha} + c_U\delta\right\} \text{ mit } c_U = \frac{1}{\sqrt{2\bar{\vartheta}(1-\bar{\vartheta})}};$$

$$\beta_n^V(\delta) \longrightarrow \beta^V(\delta) = P\{N(0,1) < -z_{1-\alpha} + c_V\delta\} \text{ mit } c_V = \frac{1}{\sqrt{p_+}}.$$

Somit gilt

$$are(V:U) = \lim \frac{n_U}{n_V} = \frac{c_V^2}{c_U^2} = \frac{2\bar{\vartheta}(1-\bar{\vartheta})}{p_+}.$$

Es gilt: $\frac{1}{2} \le are(V:U) < \infty$.

*Begründung:*
Es gilt

$$2\bar{\vartheta}(1-\bar{\vartheta}) = 2\left(p_{11} + \tfrac{1}{2}p_+\right)\left(p_{22} + \tfrac{1}{2}p_+\right) = 2p_{11}p_{22} + p_+\underbrace{(p_{11} + p_{22})}_{=1-p_+} + \tfrac{1}{2}p_+^2$$

$$= 2p_{11}p_{22} + p_+ - \tfrac{1}{2}p_+^2,$$

und somit

$$\frac{2\bar{\vartheta}(1-\bar{\vartheta})}{p_+} = \frac{2p_{11}p_{22}}{p_+} + 1 - \tfrac{1}{2}p_+ \longrightarrow \infty \text{ für } p_+ \to 0 \ (p_{11}, p_{22} > 0),$$

d.h. es gilt $are(V:U) \to \infty$ für $p_+ \to 0$ $(p_{11}, p_{22} > 0)$. Andererseits haben wir

$$\frac{2\bar{\vartheta}(1-\bar{\vartheta})}{p_+} = \frac{2p_{11}p_{22}}{p_+} + 1 - \tfrac{1}{2}p_+ \ge 1 - \tfrac{1}{2}p_+ \ge \tfrac{1}{2} \text{ (denn } p_+ \le 1),$$

und daher gilt $are(V:U) \ge \frac{1}{2}$.

*Bemerkung:*
Die beiden Grenzfälle der asymtotischen relativen Effizienz sind recht plausibel. Wenn $p_{11} = p_{22} = 0$ und damit $p_+ = 1$, so wechseln alle Versuchspersonen bei der Behandlung ihr Befinden (positiv wird negativ und negativ wird positiv). Es ist dann nicht sinnvoll, die gleichen Versuchspersonen zweimal zu untersuchen (vorher und nachher), die zweite Untersuchung ist also beim Test von McNemar völlig wertlos. Beim Vier-Felder-Test werden zwei unabhängige Stichproben je mit Umfang $n = 100$ (beispielsweise) gezogen, und damit steht diesem Test die doppelte Informationsmenge (d.h. eine doppelt so große Stichprobe) zur Verfügung im Vergleich zum Test von McNemar, und daher ist es plausibel, dass $are(V:U) = \frac{1}{2}$.

Andererseits gilt $p_{11} + p_{22} \to 1$ für $p_+ \to 0$, d.h. die meisten Versuchspersonen ändern dann ihr Befinden nicht bei der Behandlung. Dann ist der Versuchsplan mit verbundenen Stichproben sinnvoller als der Plan mit unabhängigen Stichproben; beim zweiten Versuchsplan verwischen die zahlreichen Versuchspersonen, welche bei der Behandlung unbeeinflusst bleiben, die (geringen) Unterschiede zwischen $p_{12}$ und $p_{21}$; der Test von McNemar kann diese Unterschiede besser erkennen, da die irrelevanten Fälle $X_{11}$ und $X_{22}$ unberücksichtigt bleiben können. Die Situation ist hier ähnlich wie beim Vorzeichentest, wo es auch sinnvoll war, die indifferenten Fälle unberücksichtigt zu lassen. Als grobe Faustregel haben wir also:

> *U* besser als *V*, falls Hauptdiagonale der Vier-Felder-Tafel schwach besetzt, d.h. $X_{11}, X_{22}$ klein im Vergleich zu $X_{12}, X_{21}$;
>
> *V* besser als *U*, falls Hauptdiagonale der Vier-Felder-Tafel stark besetzt, d.h. $X_{11}, X_{22}$ groß im Vergleich zu $X_{12}, X_{21}$.

*Beispiele:*

a) $\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \cdot \\ \cdot & \frac{1}{4} \end{pmatrix}$

Hier ist: $\left.\begin{aligned} p_+ &= p_{12} + p_{21} = \tfrac{1}{2} \\ \bar{\vartheta} &= p_{11} + \tfrac{1}{2} p_+ = \tfrac{1}{2} \end{aligned}\right\} \Rightarrow \; are(V:U) = \dfrac{1/2}{1/2} = 1.$

b) $\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.45 & \cdot \\ \cdot & 0.45 \end{pmatrix}$

Hier ist: $\left.\begin{aligned} p_+ &= p_{12} + p_{21} = 0.1 \\ \bar{\vartheta} &= p_{11} + \tfrac{1}{2} p_+ = \tfrac{1}{2} \end{aligned}\right\} \Rightarrow \; are(V:U) = \dfrac{0.5}{0.1} = 5.$

c) $\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0 & \cdot \\ \cdot & 0 \end{pmatrix}$

Hier ist: $\left.\begin{aligned} p_+ &= p_{12} + p_{21} = 1 \\ \bar{\vartheta} &= p_{11} + \tfrac{1}{2} p_+ = \tfrac{1}{2} \end{aligned}\right\} \Rightarrow \; are(V:U) = \dfrac{1/2}{1} = \tfrac{1}{2}.$

## 5. Vergleich bei endlichem Stichprobenumfang

Wir wollen hier die beiden Versuchspläne bei endlichem Stichprobenumfang vergleichen.

*A) Exakter Vier-Felder-Test von Fisher (unabhängige Stichproben)*

Gemäß (2) ist die Gütefunktion gegeben durch

$$\begin{aligned}\beta_n^U(\vartheta_1,\vartheta_2) &= \sum_{k=0}^{2n} P\{X=k\}\,P\{\text{Test verwift } H_0 \,|\, X=k\} = \sum_{k=0}^{2n} P\{X=k\}\,P\{X_1>k_r \,|\, X=k\} \\ &= \sum_{k=0}^{2n} P\{X_1>k_r,\, X=k\} = \sum_{k=0}^{2n}\left(\sum_{j=k_r+1}^{\min(n,k)} P\{X_1=j,\, X_2=k-j\}\right).\end{aligned} \tag{10}$$

Dabei sind die beiden Zufallsgrößen $X_1$ und $X_2$ unabhängig und besitzen je eine Binomialverteilung $Bi(n,\vartheta_1)$ bzw. $Bi(n,\vartheta_2)$; $k_r$ ist die rechte kritische $\alpha$-Schranke zur hypergeometrischen Verteilung $H(2n,n,k)$. Ein Minitab-Makro zur Berechnung der Gütefunktion findet sich unten.

*B) Test von McNemar (verbundene Stichproben)*

Gemäß (4) ist die Gütefunktion gegeben durch

$$\begin{aligned}\beta_n^V(p_{12},p_{21}) &= P\{\text{Test verwirft } H_0\} = \sum_{m=0}^{n} P\{X_+=m\}\,P\{\text{Test verwift } H_0 \,|\, X_+=m\} \\ &= \sum_{m=0}^{n} P\{Bi(n,p_+)=m\}\,P\{Bi(m,\eta)>k_r\} \quad \text{mit} \quad \eta=\frac{p_{12}}{p_+}.\end{aligned}$$

Dabei ist $X_+ = X_{12}+X_{21} \sim Bi(n,p_+)$ mit $p_+ = p_{12}+p_{21}$, und $k_r$ ist die rechte kritische $\alpha$-Schranke zur Binomialverteilung $Bi(m,\frac{1}{2})$. Ein Minitab-Makro zur Berechnung der Gütefunktion findet sich unten.

*Vergleich der Gütefunktionen*

Falls $p_{11}=p_{22}=\frac{1}{4}$, so ist $p_+ = 1-p_{11}-p_{22}=\frac{1}{2}$ und die asymptotische relative Effizienz zwischen den beiden Versuchsplänen beträgt $are(V:U)=1$, wie wir oben im Beispiel a) gesehen haben. Die Gütefunktionen konvergieren also bei beiden Versuchsplänen gegen die gleiche Grenzfunktion

$$\beta_n^*(\delta) \xrightarrow[n\to\infty]{} \beta(\delta) = P\{N(0,1) < -z_{1-\alpha}+\delta\sqrt{2}\}.$$

Die Tabellen 1a bis 1d geben für für $n=25,\,400,\,1600$ und für ausgewählte Werte von $\delta = d\sqrt{n} = \sqrt{n}(p_{12}-p_{21})$ die Werte der exakten Gütefunktionen, sowie die Werte der asymptotischen Gütefunktion. Da die beiden Prüfpläne bei endlichem Stichprobenumfang das vorgegebene Signifikanzniveau nicht voll ausschöpfen können (diskrete Verteilungen!), so ist die Konvergenz als Funktion von *n* nicht immer monoton. Die Grenzfunktion $\beta(\delta)$ liefert einerseits schon für $n=25$ sinnvolle Näherungswerte für die exakten Gütefunktionen, andererseits aber stimmen die exakten Gütefunktionen mit der asymptotischen Gütefunktion selbst für $n=1600$ nur in etwa zwei Dezimalstellen überein. Möchte man eine Übereinstimmung in drei Dezimalstellen haben, so müßte man den Stichprobenumfang hundert mal größer wählen, also etwa $n=160\,000$, denn aufgrund unserer Tabellen sieht man, dass sich die maximale Abweichung von der Grenzfunktion etwa halbiert, wenn der Stichprobenumfang vervierfacht wird; die Konvergenzgeschwindigkeit ist also gegeben durch den Faktor $1/\sqrt{n}$.

**Tabelle 1a:** Gütefunktionen für $n = 25$, $p_{11} = p_{22} = \frac{1}{4}$, $p_+ = p_{12} + p_{21} = \frac{1}{2}$, $\alpha = 0.05$, $\delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $p_{12}$ | $p_{21}$ | $d$ | $\delta$ | $\vartheta_1$ | $\vartheta_2$ | $\eta$ | $\beta_n^U(\delta)$ | $\beta_n^V(\delta)$ | $\beta(\delta)$ |
| 0.25 | 0.25 | 0 | 0 | 0.5 | 0.5 | 0.5 | 0.0325 | 0.0277 | 0.0500 |
| 0.3 | 0.2 | 0.1 | 0.5 | 0.55 | 0.45 | 0.6 | 0.1274 | 0.1115 | 0.1742 |
| 0.35 | 0.15 | 0.2 | 1 | 0.6 | 0.4 | 0.7 | 0.3357 | 0.3098 | 0.4088 |
| 0.4 | 0.1 | 0.3 | 1.5 | 0.65 | 0.35 | 0.8 | 0.6216 | 0.6161 | 0.6831 |
| 0.45 | 0.05 | 0.4 | 2 | 0.7 | 0.3 | 0.9 | 0.8595 | 0.8960 | 0.8817 |
| 0.5 | 0 | 0.5 | 2.5 | 0.75 | 0.25 | 1 | 0.9713 | 0.9995 | 0.9707 |

**Tabelle 1b:** Gütefunktionen für $n = 100$, $p_{11} = p_{22} = \frac{1}{4}$, $p_+ = p_{12} + p_{21} = \frac{1}{2}$, $\alpha = 0.05$, $\delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $p_{12}$ | $p_{21}$ | $d$ | $\delta$ | $\vartheta_1$ | $\vartheta_2$ | $\eta$ | $\beta_n^U(\delta)$ | $\beta_n^V(\delta)$ | $\beta(\delta)$ |
| 0.25 | 0.25 | 0 | 0 | 0.5 | 0.5 | 0.5 | 0.0384 | 0.0377 | 0.0500 |
| 0.275 | 0.225 | 0.05 | 0.5 | 0.525 | 0.475 | 0.55 | 0.1441 | 0.1416 | 0.1742 |
| 0.3 | 0.2 | 0.1 | 1 | 0.55 | 0.45 | 0.6 | 0.3620 | 0.3589 | 0.4088 |
| 0.325 | 0.175 | 0.15 | 1.5 | 0.575 | 0.425 | 0.65 | 0.6408 | 0.6429 | 0.6831 |
| 0.35 | 0.15 | 0.2 | 2 | 0.6 | 0.4 | 0.7 | 0.8603 | 0.8686 | 0.8817 |
| 0.375 | 0.125 | 0.25 | 2.5 | 0.625 | 0.375 | 0.75 | 0.9651 | 0.9722 | 0.9707 |
| 0.4 | 0.1 | 0.3 | 3 | 0.65 | 0.35 | 0.8 | 0.9948 | 0.9972 | 0.9953 |
| 0.425 | 0.075 | 0.35 | 3.5 | 0.675 | 0.325 | 0.85 | 0.9996 | 0.9999 | 0.9995 |
| 0.45 | 0.05 | 0.4 | 4 | 0.7 | 0.3 | 0.9 | 1.0000 | 1.0000 | 1.0000 |

**Tabelle 1c:** Gütefunktionen für $n = 400$, $p_{11} = p_{22} = \frac{1}{4}$, $p_+ = p_{12} + p_{21} = \frac{1}{2}$, $\alpha = 0.05$, $\delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $p_{12}$ | $p_{21}$ | $d$ | $\delta$ | $\vartheta_1$ | $\vartheta_2$ | $\eta$ | $\beta_n^U(\delta)$ | $\beta_n^V(\delta)$ | $\beta(\delta)$ |
| 0.25 | 0.25 | 0 | 0 | 0.5 | 0.5 | 0.5 | 0.0416 | 0.0431 | 0.0500 |
| 0.2625 | 0.2375 | 0.025 | 0.5 | 0.5125 | 0.4875 | 0.525 | 0.1525 | 0.1566 | 0.1742 |
| 0.275 | 0.225 | 0.05 | 1 | 0.525 | 0.475 | 0.55 | 0.3752 | 0.3819 | 0.4088 |
| 0.2875 | 0.2125 | 0.075 | 1.5 | 0.5375 | 0.4625 | 0.575 | 0.6520 | 0.6597 | 0.6831 |
| 0.3 | 0.2 | 0.1 | 2 | 0.55 | 0.45 | 0.6 | 0.8646 | 0.8707 | 0.8817 |
| 0.3125 | 0.1875 | 0.125 | 2.5 | 0.5625 | 0.4375 | 0.625 | 0.9652 | 0.9683 | 0.9707 |
| 0.325 | 0.175 | 0.15 | 3 | 0.575 | 0.425 | 0.65 | 0.9943 | 0.9952 | 0.9953 |
| 0.3375 | 0.1625 | 0.175 | 3.5 | 0.5875 | 0.4125 | 0.675 | 0.9994 | 0.9996 | 0.9995 |
| 0.35 | 0.15 | 0.2 | 4 | 0.6 | 0.4 | 0.7 | 1.0000 | 1.0000 | 1.0000 |

**Tabelle 1d:** Gütefunktionen für $n = 1600$, $p_{11} = p_{22} = \frac{1}{4}$, $p_+ = p_{12} + p_{21} = \frac{1}{2}$, $\alpha = 0.05$, $\delta = d\sqrt{n}$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| $p_{12}$ | $p_{21}$ | $d$ | $\delta$ | $\vartheta_1$ | $\vartheta_2$ | $\eta$ | $\beta_n^U(\delta)$ | $\beta_n^V(\delta)$ | $\beta(\delta)$ |
| 0.25 | 0.25 | 0 | 0 | 0.5 | 0.5 | 0.5 | 0.0465 | 0.0465 | 0.0500 |
| 0.25625 | 0.24375 | 0.0125 | 0.5 | 0.50625 | 0.49375 | 0.5125 | 0.1654 | 0.1653 | 0.1742 |
| 0.2625 | 0.2375 | 0.025 | 1 | 0.5125 | 0.4875 | 0.525 | 0.3955 | 0.3953 | 0.4088 |
| 0.26875 | 0.23125 | 0.0375 | 1.5 | 0.51875 | 0.48125 | 0.5375 | 0.6709 | 0.6711 | 0.6831 |
| 0.275 | 0.225 | 0.05 | 2 | 0.525 | 0.475 | 0.55 | 0.8750 | 0.8755 | 0.8817 |
| 0.28125 | 0.21875 | 0.0625 | 2.5 | 0.53125 | 0.46875 | 0.5625 | 0.9685 | 0.9689 | 0.9707 |
| 0.2875 | 0.2125 | 0.075 | 3 | 0.5375 | 0.4625 | 0.575 | 0.9949 | 0.9950 | 0.9953 |
| 0.29375 | 0.20625 | 0.0875 | 3.5 | 0.54375 | 0.45625 | 0.5875 | 0.9995 | 0.9995 | 0.9995 |
| 0.3 | 0.2 | 0.1 | 4 | 0.55 | 0.45 | 0.6 | 1.0000 | 1.0000 | 1.0000 |

Minitab-Makro `minitab3.mac` zur Berechnung der Gütefunktion $\beta_n^U(\vartheta_1,\vartheta_2)$

```
gmacro
note
# Exakte Berechung der Gütefunktion beim Vier-Felder-Test
# mit gleichen Stichprobenumfängen: n1=n2=n
# H0: theta1 <= theta2  gegen  H1: theta1 > theta2
#
erase c1-c5 k5-k12
name c1 'k-Werte' c2 'kr(2n,n,k,alpha)' c3 'prob_k'
name k1 'n' k2 'alpha' k3 'theta1' k4 'theta2' k5 'nn' k6 'term1' k7 'term2'
name k8 'k' k9 'kr' k10 'summe' k11 'probk' k12 'alphac' k13 'j' k14 'jc'
name k15 'krplus1' k16 'jend'
note
note  Eingabewerte: K1 = n (Stichprobenumfang)
note                K2 = alpha = Signifikanzniveau
note                K3 = theta1
note                K4 = theta2
note
print k1-k4
note Sind die Eingabewerte korrekt? (y/n)?
yesno k10
if k10 = 0
  exit
endif

let 'nn'  = 2*'n'
do 'k'=1:'nn'
  # für k=0 ist die Ablehnw'keit stets =0
  # kritische Schranke zu H(2n,n,k)
  let 'alphac' = 1-'alpha'
  InvCDF 'alphac' 'kr';
    Hypergeo 'nn' 'n' 'k'.
  let 'krplus1' = 'kr'+1
  # Berechnung von P{X1>kr, X=k}
  if 'krplus1' > 'k' or 'krplus1' > 'n'
    let 'summe' = 0
  else
    let 'summe' = 0
    if 'n' < 'k'
      let 'jend' = 'n'
    else
      let 'jend' = 'k'
    endif
    do 'j'='krplus1':'jend'
      PDF 'j' 'term1';
        Binomial 'n' 'theta1'.
      let 'jc' = 'k'-'j'
      PDF 'jc' 'term2';
        Binomial 'n' 'theta2'.
      let 'summe' = 'summe' + 'term1'*'term2'
    enddo
  endif
  let c1('k') = 'k'
  let c2('k') = 'kr'
  let c3('k') = 'summe'
enddo
let 'summe' = sum(c3)
print 'summe'
note summe = Ablehnwahrscheinlichkeit (Wert der Gütefunktion)
endmacro
```

Minitab-Makro `minitab4.mac` zur Berechnung der Gütefunktion $\beta_n^V(\vartheta_1, \vartheta_2)$

```
gmacro
note
# Exakte Berechung der Gütefunktion beim Test von McNemar
# H0: p12 = p21  gegen  H1: p12 > p21
#
erase c1-c5 k5-k12
name c1 'm-Werte' c2 'kr(m,alpha)' c3 'prob_1' c4 'prob_2' c5 'product'
name k1 'n' k2 'alpha' k3 'p12' k4 'p21' k5 'p_plus' k6 'eta'
name k7 'm' k8 'kr' k9 'summe' k10 'prob1' k11 'prob2' k12 'alphac'
note
note  Eingabewerte: K1 = n (Stichprobenumfang)
note                K2 = alpha = Signifikanzniveau
note                K3 = p12
note                K4 = p21
note
prin k1-k4
note Sind die Eingabewerte korrekt? (y/n)?
yesno k10
if k10 = 0
  exit
endif

let 'p_plus'  = 'p12'+'p21'
let 'eta' = 'p12'/'p_plus'

let 'summe' = 0
do 'm'=1:'n'
#do 'm'=650:950
  # für m=0 ist die Ablehnw'keit stets =0
  PDF 'm' 'prob1';
    Binomial 'n' 'p_plus'.
  # kritische Schranke zu Bi(m,1/2)
  let 'alphac' = 1-'alpha'
  InvCDF 'alphac' 'kr';
    Binomial 'm' 0.5.
  CDF 'kr' 'prob2';
    Binomial 'm' 'eta'.
  let 'prob2' = 1-'prob2'
  let 'summe' = 'summe'+'prob1'*'prob2'
  let c1('m') = 'm'
  let c2('m') = 'kr'
  let c3('m') = 'prob1'
  let c4('m') = 'prob2'
  let c5('m') = 'prob1'*'prob2'
enddo
prin summe
note summe = Ablehnwahrscheinlichkeit (Wert der Gütefunktion)
endmacro
```

# 7. Lineare Rangtests beim Zweistichprobenproblem

In diesem Kapitel wollen wir lineare Rangtests zum Zweistichprobenproblem untersuchen, welche als Konkurrenten zum klassischen doppelten t-Test in Frage kommen. Dazu benötigen wir zunächst einige Grundlagen über *Ordnungs-* und *Rangstatistiken*.

## 1. Ordnungsstatistiken und Rangstatistiken (order and rank statistics)

Lit.: Büning-Trenkler (1994), Kap. 3, 4.

$X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit gleicher Verteilung;
$x_1,\ldots,x_n$ zugehörige Realisationen;
$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ geordnete Stichprobe;

*Ordnungsstatistiken (order statistics)*

Die Zufallsgößen $X_{(k)}$ seien so definiert, dass $x_{(k)} =$ Realisation von $X_{(k)}$ für $k = 1,\ldots,n$. Dann gilt $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$. Die Zufallsgrößen $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ heißen *Ordnungsstatistiken (order statistics)* zur Zufallsstichprobe $X_1,\ldots,X_n$. Es gilt $X_{(1)} = \min(X_1,\ldots,X_n)$, $X_{(n)} = \max(X_1,\ldots,X_n)$; Zentralwert, Quartile, Spannweite, etc. sind Ordnungsstatistiken bzw. Funktionen von Ordnungsstatistiken

*Rangstatistiken (Rangzahlen, rank statistics)*

Wir wollen annehmen, dass $x_1,\ldots,x_n$ alle verschieden sind (keine Bindungen, no ties). Dann sind die Rangzahlen zu $x_1,\ldots,x_n$ wie folgt definiert:

$r_i = rang(x_i) =$ Platznummer von $x_i$ in der geordneten Liste $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$

$$= \#\left\{ j \,|\, x_j \le x_i \right\} = \sum_{j=1}^{n} \chi(x_i - x_j), \quad \text{wobei } \chi(t) = \begin{cases} 1 \text{ für } t \ge 0 \\ 0 \text{ für } t < 0 \end{cases}$$

$R_i = rang(X_i) =$ Zufallsgröße mit Realisation $r_i$.

*Beispiel:*

| $i$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $x_i$ | 4 | 17 | 3 | 9 |
| $rang(x_i)$ | 2 | 4 | 1 | 3 |

*Zur Verteilung der Rangzahlen*

$X_1,\ldots,X_n$ unabhängig, je mit gleicher *stetiger* Verteilung;
es gilt: $P\left\{X_i = X_j\right\} = 0$ für $i \ne j$, d.h. die Wahrscheinlichkeit, dass Bindungen auftreten ist Null;
$R_1,\ldots,R_n$ zugehörige Rangzahlen (als Zufallsgrößen).

1. Gemeinsame Verteilung von $(R_1,\ldots,R_n)$:
   $P\left\{R_1 = r_1,\ldots,R_n = r_n\right\} = 1/n!$ wobei $r_1,\ldots,r_n$ eine beliebige Permutation von $1,\ldots,n$ ist.
   Begründung: Symmetrie!
2. Eindimensionale Verteilung: $R_1,\ldots,R_n$ besitzen je die gleiche Verteilung, und zwar:
   $P\left\{R_1 = k\right\} = 1/n$ für $k = 1,\ldots,n$.
3. Zweidimensionale Verteilung: Die Paare $(R_i, R_j)$ mit $i \ne j$ besitzen je die gleiche Verteilung, und zwar: $P\left\{R_1 = k, R_2 = \ell\right\} = \dfrac{1}{n(n-1)}$ für $k, \ell = 1,\ldots,n, k \ne \ell$.

4. $E(R_1)=\sum_{k=1}^{n} k\,P\{R_1=k\}=\frac{1}{n}(1+\cdots+n)=\frac{n+1}{2}$; somit gilt $E(R_i)=\frac{n+1}{2}$ für $i=1,\ldots,n$.

5. $Var(R_1)=E(R_1^2)-\big[E(R_1)\big]^2$; $E(R_1^2)=\frac{1}{n}\sum_{k=1}^{n}k^2=\frac{1}{6}(n+1)(2n+1)$

$$\Rightarrow Var(R_1)=\tfrac{1}{6}(n+1)(2n+1)-\frac{(n+1)^2}{4}=\frac{n+1}{12}\underbrace{\big[2(2n+1)-3(n+1)\big]}_{=n-1}=\frac{(n+1)(n-1)}{12}=\frac{n^2-1}{12};$$

somit gilt: $Var(R_i)=\frac{n^2-1}{12}$ für $i=1,\ldots,n$.

6. $Cov(R_1,R_2)=E(R_1R_2)-E(R_1)E(R_2)=E(R_1R_2)-\frac{(n+1)^2}{4}$;

$$E(R_1R_2)=\sum_{k\neq\ell} k\ell\,P\{R_1=k,R_2=\ell\}=\frac{1}{n(n-1)}\left(\sum_{k,\ell}k\ell-\sum_k k^2\right)$$

$$=\frac{1}{n(n-1)}\left(\frac{n^2(n+1)^2}{4}-\frac{n(n+1)(2n+1)}{6}\right)=\frac{n+1}{12(n-1)}\underbrace{\big[3n(n+1)-2(2n+1)\big]}_{=3n^2-n-2}$$

$$=\frac{n+1}{12(n-1)}(3n^2-n-2);$$

$$Cov(R_1,R_2)=\frac{n+1}{12(n-1)}(3n^2-n-2)-\frac{(n+1)^2}{4}$$

$$=\frac{n+1}{12(n-1)}\underbrace{\big[(3n^2-n-2)-3(n+1)(n-1)\big]}_{=-(n-1)}=-\frac{n+1}{12};$$

$$Corr(R_1,R_2)=\frac{Cov(R_1,R_2)}{Var(R_1)}=-\frac{(n+1)/12}{(n+1)(n-1)/12}=-\frac{1}{n-1};$$

somit gilt: $Cov(R_i,R_j)=-\frac{n+1}{12}$ und $Corr(R_i,R_j)=-\frac{1}{n-1}$ für $i,j=1,\ldots,n,\ i\neq j$.

## 2. Klassischer doppelter t-Test

Wir betrachten hier das klassische Zweistichprobenproblem.

$\left.\begin{matrix}X_1,\ldots,X_m\\ Y_1,\ldots,Y_n\end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases}\text{je mit } N(\mu_1,\sigma^2)\\ \text{je mit } N(\mu_2,\sigma^2);\end{cases}$

$\mu_1,\mu_2,\sigma$ unbekannt;

$H_0: \mu_1=\mu_2$ (beide Stichproben haben gleiche Verteilung);

$H_1: \mu_1<\mu_2$ (*Y*-Werte im Mittel größer als *X*-Werte);

klassischer t-Test:

$$T=\frac{\bar Y-\bar X}{S}\sqrt{\frac{mn}{m+n}},\ \text{wobei } S^2=\frac{1}{m+n-2}\Big(\sum(X_i-\bar X)^2+\sum(Y_j-\bar Y)^2\Big);$$

falls $\mu_1=\mu_2$: $T\sim t(m+n-2)$;

Testvorschrift: $H_0$ ablehnen, falls $T>t_{1-\alpha}(m+n-2)$.

Aufgrund des zentralen Grenzwertsatzes kann der t-Test auch angewandt werden, wenn keine Normalverteilung vorliegt, falls die Stichprobenumfänge genügend groß sind.

## 3. Lineare Rangtests beim Lagemodell

Jetzt betrachten wir das Zweistichprobenproblem in etwas allgemeinerer Form als beim t-Test.

$$\left.\begin{matrix} X_1,\ldots,X_m \\ Y_1,\ldots,Y_n \end{matrix}\right\} \text{ unabhängige Zufallsgrößen } \begin{cases} \text{je mit Dichte } f(x-\mu_1) \\ \text{je mit Dichte } f(x-\mu_2); \end{cases}$$

dabei sei *f* eine beliebige Dichtefunktion mit $f(x)>0$ für $x\in\mathbb{R}$, und wir interessieren uns für das gleiche Testproblem wie oben. Wir können annehmen, dass $\mu_1=0$, und wir setzen dann $\mu_2=\vartheta$. Dann lautet das Testproblem

$$\left.\begin{matrix} X_1,\ldots,X_m \\ Y_1,\ldots,Y_n \end{matrix}\right\} \text{ unabhängige Zufallsgrößen } \begin{cases} \text{je mit Dichte } f(x) \\ \text{je mit Dichte } f(x-\vartheta); \end{cases}$$

$H_0: \vartheta = 0$ gegen $H_1: \vartheta > 0$.

*Test von Wilcoxon-Mann-Whitney*

Gesamtstichprobe $X_1,\ldots,X_m,Y_1,\ldots Y_n$ ordnen (keine Bindungen, da stetige Verteilung);

$$S_X = \sum_{i=1}^{m} rang(X_i) = \text{Rangsumme der } X\text{-Beobachtungen;}$$

$$S_Y = \sum_{j=1}^{n} rang(Y_j) = \text{Rangsumme der } Y\text{-Beobachtungen;}$$

es gilt: $S_X + S_Y = 1+\cdots+N = \dfrac{N(N+1)}{2}$, wobei $N=m+n$;

unter $H_0$ haben die $X_i$ und die $Y_j$ alle die gleiche Dichte $f(x)$, und daher ist die exakte Verteilung von $S_X$ und $S_Y$ dann unabhängig von der Dichte $f(x)$ (*verteilungsfreies* Verfahren). Die exakte Verteilung ist berechenbar mit kombinatorischen Verfahren (vgl. StatXact). Die Nullhypothese wird abgelehnt, falls die *Y*-Werte zu groß sind, d.h. falls $S_Y$ zu groß ist (bzw. $S_X$ zu klein). Der kritische Wert wir bestimmt aufgrund der exakten Verteilung von $S_Y$ unter der Nullhypothese. Äquivalent dazu ist die Methode mit dem p-Wert:

Es sei $s_{\text{obs}} = \sum_{j=1}^{n} rang(y_j) =$ Rangsumme für konkret vorliegende Stichprobe (obs = observed);

berechne $p_{crit} = P_0\{S_Y \geq s_{\text{obs}}\}$, wobei $P_0\{\cdots\}$ die exakte Verteilung unter $H_0$ bezeichnet;

verwirf $H_0$, falls $p_{crit} \leq \alpha$.

Dies ist ein Test zum Signifikanzniveau $\alpha$. Der Test ist verteilungsfrei, da keinerlei Annahme über die Dichte *f* gemacht werden müssen. Der Test geht zurück auch Wilcoxon-Mann-Whitney. Wir werden später sehen, dass der Test so modifiziert werden kann, dass er auch im unstetigen Fall korrekt angewandt werden kann.

*Allgemeine lineare Rangtests*

Gesamtstichprobe $X_1,\ldots,X_m,Y_1,\ldots Y_n$ ordnen (keine Bindungen, da stetige Verteilung);

$w_1,\ldots,w_N$ Gewichte, $N=m+n$, $w_1<w_2<\cdots<w_N$, d.h. große Beobachtungen $\rightarrow$ großes Gewicht;

die Forderung $w_1<\cdots<w_N$ kann abgeschwächt werden (vgl. unten);

im Spezialfall des Tests von Wilcoxon gilt: $w_i=i,\ i=1,\ldots,N$;

Gewicht von $x_i$: falls $rang(x_i)=r_i=7$ so ist $gew(x_i)=w_{r_i}=w_7$;

Gewicht von $X_i$: falls $R_i=rang(X_i)$, so ist $gew(X_i)=w_{R_i}$;

Gewicht von $Y_j$: falls $R_j=rang(Y_j)$, so ist $gew(Y_j)=w_{R_j}$;

$S_X = \sum_{i=1}^{m} gew(X_i)$ = Gewichtssumme der $X$-Beobachtungen;

$S_Y = \sum_{j=1}^{n} gew(Y_j)$ = Gewichtssumme der $Y$-Beobachtungen;

es gilt: $S_X + S_Y = w_1 + \cdots + w_N$ ;

unter $H_0$ haben die $X_i$ und die $Y_j$ alle die gleiche Dichte $f(x)$, und daher ist die exakte Verteilung von $S_X$ und $S_Y$ dann unabhängig von der Dichte $f(x)$ (*verteilungsfreies* Verfahren). Die exakte Verteilung ist berechenbar mit kombinatorischen Verfahren (vgl. StatXact). Die Nullhypothese wird abgelehnt, falls die *Y*-Werte zu groß sind, d.h. falls $S_Y$ zu groß ist (bzw. $S_X$ zu klein). Der kritische Wert wir bestimmt aufgrund der exakten Verteilung von $S_Y$ unter der Nullhypothese. Äquivalent dazu ist die Methode mit dem p-Wert:

Es sei $s_{\text{obs}} = \sum_{j=1}^{n} gew(y_j) =$ Gewichtssumme für konkret vorliegende Stichprobe;

berechne $p_{crit} = P_0\{S_Y \geq s_{\text{obs}}\}$, wobei $P_0\{\cdots\}$ die exakte Verteilung unter $H_0$ bezeichnet;

verwirf $H_0$, falls $p_{crit} \leq \alpha$ .

Dies ist ein Test zum Signifikanzniveau $\alpha$ . Der Test ist verteilungsfrei, da keinerlei Annahme über die Dichte *f* gemacht werden müssen.

*Zur Bestimmung der exakten Verteilung von $S_Y$ unter der Nullhypothese:*

$X_i, Y_j$ unabhängig je mit gleicher stetiger Verteilungsfunktion $F$ ;

$\Rightarrow$ $P\{X_1 < X_2 < \cdots < X_m < Y_1 < \cdots < Y_n\} = 1/N!$, wobei $N = m + n$ ;
das gleiche gilt für jede andere Permutation der Variablen $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ ;

$\Rightarrow$ $P\{X < X < \cdots < X < Y < Y < \cdots < Y\} = 1 \Big/ \binom{m+n}{m} = \frac{m!n!}{(m+n)!}$ ;

das gleiche gilt für jede andere Kombination der *X*- und *Y*-Variablen.

*Beispiel:* $m = 3, n = 2 : X_1, X_2, X_3, Y_1, Y_2$ ; $w_i = i$ (Rangzahlen als Gewichte)

Es gibt $\binom{5}{2} = 10$ unterscheidbare Kombinationen von *X*- und *Y*-Variablen;

jede Kombination hat die Wahrscheinlichkeit $1/10$ :

| | Ränge (= Gewichte) 1 | 2 | 3 | 4 | 5 | $S_Y$ |
|---|---|---|---|---|---|---|
| 1 | *X* | *X* | *X* | *Y* | *Y* | 9 |
| 2 | *X* | *X* | *Y* | *X* | *Y* | 8 |
| 3 | *X* | *X* | *Y* | *Y* | *X* | 7 |
| 4 | *X* | *Y* | *X* | *X* | *Y* | 7 |
| 5 | *X* | *Y* | *X* | *Y* | *X* | 6 |
| 6 | *X* | *Y* | *Y* | *X* | *X* | 5 |
| 7 | *Y* | *X* | *X* | *X* | *Y* | 6 |
| 8 | *Y* | *X* | *X* | *Y* | *X* | 5 |
| 9 | *Y* | *X* | *Y* | *X* | *X* | 4 |
| 10 | *Y* | *Y* | *X* | *X* | *X* | 3 |

Exakte Verteilung von $S_Y$

| $k$ | $P\{S_Y = k\}$ |
|---|---|
| 3 | 1/10 |
| 4 | 1/10 |
| 5 | 2/10 |
| 6 | 2/10 |
| 7 | 2/10 |
| 8 | 1/10 |
| 9 | 1/10 |
| | 1 |

Nun stellen sich folgende Fragen:

a) Vergleich des t-Tests mit dem Test von Wilcoxon;
b) Optimale Gewichtung bei gegebener Dichte $f(x)$ ;
c) Asymptotische Gütefunktionen, asymptotische relative Effizienz.

## 4. Asymptotische Gütefunktion beim t-Test

$\left.\begin{matrix} X_1,\ldots,X_m \\ Y_1,\ldots,Y_n \end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases} \text{je mit Dichte } f(x) \\ \text{je mit Dichte } f(x-\vartheta); \end{cases}$

$\sigma^2 = Var(X_i) = Var(Y_j),\ 0 < \sigma^2 < \infty$;

$H_0: \vartheta = 0$ gegen $H_1: \vartheta > 0$.

*t-Test:*

$$T = \frac{\bar{Y} - \bar{X}}{S}\sqrt{\frac{mn}{m+n}},\ \text{wobei}\ S^2 = \frac{1}{m+n-2}\left(\sum (X_i - \bar{X})^2 + \sum (Y_j - \bar{Y})^2\right);$$

falls $\vartheta = 0$ und $X_i, Y_j$ normalverteilt: $T \sim t(m+n-2)$;

falls $\vartheta = 0$, $m,n$ groß, und $0 < \sigma^2 < \infty$: $T \approx N(0,1)$ (aufgrund des zentralen Grenzwertsatzes);

Testvorschrift: $H_0$ ablehnen, falls $T > t_{1-\alpha}(m+n-2)$ bzw. falls $T > z_{1-\alpha}$.

Wir betrachten den folgenden Grenzübergang:

$$N = m+n \to \infty;$$

(1) $\frac{m}{N} \to \lambda,\ \frac{n}{N} \to 1-\lambda,\ 0 < \lambda < 1;$

$\delta = \vartheta\sqrt{N}$ fest $\left(\Rightarrow \vartheta = \frac{\delta}{\sqrt{N}} \to 0\right)$.

*Gütefunktion*

Die Gütefunktion des t-Tests kann (bei Annahme einer Normalverteilung) geschrieben werden als

$$(2)\quad \beta_N(\delta) = P\{\text{Test verwirft } H_0\} = P\left\{\frac{\bar{Y} - \bar{X}}{S}\sqrt{\frac{mn}{N}} > t_{1-\alpha}(N-2)\right\},$$

und es gilt

$$\frac{\bar{Y} - \bar{X}}{S}\sqrt{\frac{mn}{N}} = \underbrace{\frac{\bar{Y} - \vartheta - \bar{X}}{S}\sqrt{\frac{mn}{N}}}_{\sim t(N-2)} + \frac{\vartheta}{S}\sqrt{\frac{mn}{N}}.$$

Beim Grenzübergang (1) gilt

$$S^2 = \frac{1}{N-2}\left(\sum \left(X_i - \bar{X}\right)^2 + \sum \left(Y_j - \bar{Y}\right)^2\right) \xrightarrow{P} \sigma^2;$$

$$\frac{\bar{Y} - \vartheta - \bar{X}}{S}\sqrt{\frac{mn}{N}} \sim t(N-2) \longrightarrow N(0,1)$$

$$\frac{\vartheta}{S}\sqrt{\frac{mn}{N}} = \frac{\vartheta\sqrt{N}}{S}\sqrt{\frac{mn}{N^2}} \longrightarrow \frac{\delta}{\sigma}\sqrt{\lambda(1-\lambda)}.$$

Damit gilt für die Gütefunktion beim Grenzübergang (1)

$$(3)\quad \beta_N(\delta) \longrightarrow \beta(\delta) = P\left\{N(0,1) + \frac{\delta}{\sigma}\sqrt{\lambda(1-\lambda)} > z_{1-\alpha}\right\} = P\left\{N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sigma}\sqrt{\lambda(1-\lambda)}\right\}.$$

Die Beziehung (3) gilt auch ohne Normalverteilungsannahme unter der Voraussetzung $0 < \sigma^2 < \infty$ (zentraler Grenzwertsatz).

## 5. Asymptotische Gütefunktion beim allgemeinen linearen Rangtest

Lit.: Hettmansperger (1984), Chapter 3, 132-177.

Wir wollen bei der Gütefunktion zum allgemeinen linearen Rangtest den gleichen Grenzübergang (1) wie beim t-Test durchführen. Wir betrachten das gleiche Testproblem wie beim t-Test:

$\left.\begin{matrix} X_1,\ldots,X_m \\ Y_1,\ldots,Y_n \end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases} \text{je mit Dichte } f(x) \\ \text{je mit Dichte } f(x-\vartheta); \end{cases}$

$H_0: \vartheta = 0$ gegen $H_1: \vartheta > 0$.

Gesamtstichprobe $X_1,\ldots,X_m,Y_1,\ldots Y_n$ ordnen (keine Bindungen, da stetige Verteilung);
$w_1,\ldots,w_N$ Gewichte;

Gewicht von $Y_j$: $gew(Y_j) = w_{R_j}$, wobei $R_j = rang(Y_j)$;

$S_Y = \sum_{j=1}^{n} w_{R_j} =$ Gewichtssumme der $Y$-Beobachtungen.

Wir wollen nun Erwartungswert und Varianz von $S_Y$ unter der Nullhypothese bestimmen, d.h. unter der Annahme, dass die $X$- und $Y$-Beobachtungen die gleiche stetige Verteilung besitzen. Dann sind $R_1,\ldots,R_n$ abhängige Zufallsgröße je mit der gleichen Verteilung:

$$P_0\{R_1 = i\} = \frac{1}{N},\ i = 1,\ldots,N;$$

$$E_0\left(w_{R_1}\right) = \sum_{i=1}^{N} w_i P_0\{R_1 = i\} = \frac{1}{N}\sum_{i=1}^{N} w_i = \bar{w};$$

$$Var_0\left(w_{R_1}\right) = E_0\left(w_{R_1}^2\right) - \bar{w}^2 = \sum_{i=1}^{N} w_i^2 P_0\{R_1 = i\} - \bar{w}^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} w_i^2 - \bar{w}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(w_i - \bar{w}\right)^2.$$

Weiter haben alle Paare $(R_i, R_j)$ mit $i \neq j$ die gleiche zweidimensionale Verteilung:

$$P_0\{R_1 = i, R_2 = j\} = \frac{1}{N(N-1)},\ i,j = 1,\ldots,N;\ i \neq j;$$

$$Cov_0\left(w_{R_1}, w_{R_2}\right) = E_0\left(w_{R_1} - \bar{w}\right)\left(w_{R_2} - \bar{w}\right) = \sum_{i \neq j}\left(w_i - \bar{w}\right)\left(w_j - \bar{w}\right) P_0\{R_1 = i, R_2 = j\}$$

$$= \frac{1}{N(N-1)}\sum_{i=1}^{N}\left[\left(w_i - \bar{w}\right)\sum_{j \neq i}\left(w_j - \bar{w}\right)\right] = -\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(w_i - \bar{w}\right)^2,$$

denn $\sum_j \left(w_j - \bar{w}\right) = 0$ und somit gilt $\sum_{j \neq i}\left(w_j - \bar{w}\right) = -\left(w_i - \bar{w}\right)$. Nun können wir Erwartungswert und Varianz von $S_Y$ unter der Nullhypothese berechnen:

$$E_0\left(S_Y\right) = n\,E_0\left(w_{R_1}\right) = n\,\bar{w};$$

$$Var_0\left(S_Y\right) = nVar_0\left(w_{R_1}\right) + n(n-1)Cov\left(w_{R_1}, w_{R_2}\right)$$

$$= \frac{n}{N}\sum_{i=1}^{N}\left(w_i - \bar{w}\right)^2 - \frac{n(n-1)}{N(N-1)}\sum_{i=1}^{N}\left(w_i - \bar{w}\right)^2$$

$$= \frac{mn}{N(N-1)}\sum_{i=1}^{N}\left(w_i - \bar{w}\right)^2.$$

Wir wollen nun den gleichen Grenzübergang (1) wie beim t-Test durchführen. Dazu ist es zweckmäßig, die Gewichte $w_i$ mit Hilfe einer gewichtserzeugenden Funktion $w(t)$ zu definieren:

$w:[0,1] \to \mathbb{R}$ gewichtserzeugende Funktion;

$w_i = w\left(\frac{i}{N+1}\right)$, $i = 1,\ldots,N$ (Einzelgewichte);

Gewicht von $Y_j$: $gew(Y_j) = w_{R_j} = w\left(\frac{R_j}{N+1}\right)$, wobei $R_j = rang(Y_j)$;

$S_Y = \sum_{j=1}^{n} w_{R_j} = \sum_{j=1}^{n} w\left(\frac{R_j}{N+1}\right) =$ Gewichtssumme der $Y$-Beobachtungen ;

$\bar{S}_Y = \frac{1}{N} S_Y$;

(4) $E_0\left(\bar{S}_Y\right) = \frac{1}{N} E_0\left(S_Y\right) = \frac{n}{N}\bar{w} = \frac{n}{N}\frac{1}{N}\sum_{i=1}^{N} w\left(\frac{i}{N+1}\right) \longrightarrow (1-\lambda)\tilde{w}$ wobei $\tilde{w} = \int_0^1 w(u)\,du$;

(5) $N\,Var_0\left(\bar{S}_Y\right) = \frac{1}{N} Var_0\left(S_Y\right) = \frac{mn}{N(N-1)}\frac{1}{N}\sum_{i=1}^{N}\left(w\left(\frac{i}{N+1}\right) - \bar{w}\right)^2 \longrightarrow \lambda(1-\lambda)\int_0^1 \left(w(u)-\tilde{w}\right)^2 du$.

*Testvorschrift beim allgemeinen linearen Rangtest*

$H_0$ ablehnen, falls $S_Y$ zu groß, d.h. falls $\bar{S}_Y > k_r$, wobei $k_r$ die rechte kritische $\alpha$-Schranke der Verteilung von $\bar{S}_Y$ unter $H_0$ ist.

*Gütefunktion:*

$$\beta_N(\delta) = P\left\{\text{Test verwirft } H_0\right\} = P\left\{\bar{S}_Y > k_r\right\} = P\left\{\frac{\bar{S}_Y - \mu}{\sigma} > \frac{k_r - \mu}{\sigma}\right\},$$

wobei $\mu = E(\bar{S}_Y)$ und $\sigma^2 = Var(\bar{S}_Y)$. (Im vorangehenden Abschnitt haben wir mit $\sigma^2$ die gemeinsame Varianz der Zufallsgrößen $X_i$ und $Y_j$ bezeichnet; diese Varianz kommt im vorliegenden Abschnitt nicht vor.) Es ist plausibel, dass bei unserem Grenzübergang (1) der folgende Grenzwertsatz gilt:

(6) $\frac{\bar{S}_Y - \mu}{\sigma} \longrightarrow N(0,1)$;

die Summanden von $\bar{S}_Y$ besitzen je die gleiche Verteilung, und sie werden für wachsende Stichprobenumfänge nahezu unabhängig, und daher kann der zentrale Grenzwertsatz zum Zuge kommen. Nun haben wir den Term

$$\frac{k_r - \mu}{\sigma} = \left(\frac{k_r - \mu_0}{\sigma_0} - \frac{\mu - \mu_0}{\sigma_0}\right)\frac{\sigma_0}{\sigma}$$

zu untersuchen. Es ist plausibel, dass

$\sigma_0/\sigma \longrightarrow 1$, denn $\vartheta = \delta/\sqrt{N} \to 0$;

$\left(k_r - \mu_0\right)/\sigma_0 \longrightarrow z_{1-\alpha}$ aufgrund des Grenzwertsatzes (6),

und es bleibt die Bestimmung des Grenzwertes von $(\mu - \mu_0)/\sigma_0$. Wir schreiben

$$\frac{\mu - \mu_0}{\sigma_0} = \frac{\vartheta}{\sigma_0} \times \frac{\mu - \mu_0}{\vartheta} = T_1 \times T_2 .$$

Beim Grenzübergang (1) gilt aufgrund von (5)

$$T_1 = \frac{\vartheta}{\sigma_0} = \frac{\delta}{\sigma_0\sqrt{N}} \longrightarrow \frac{\delta}{\tau\sqrt{\lambda(1-\lambda)}} \quad \text{mit} \quad \tau^2 = \int_0^1 \left(w(u)-\tilde{w}\right)^2 du ,$$

und im Anhang A3 wird gezeigt, dass

$$T_2 = \frac{\mu - \mu_0}{\vartheta} \longrightarrow \lambda(1-\lambda)\, I_f \,,$$

wobei

$$\begin{aligned} I_f &= \int_{-\infty}^{\infty} w'\big[F(x)\big]\, f^2(x)\, dx = -\int_{-\infty}^{\infty} w\big[F(x)\big]\, f'(x)\, dx \\ &= \int_0^1 w'(u)\, f\big[F^{-1}(u)\big]\, du = \int_0^1 w(u)\, w_f(u)\, du \quad \text{mit} \quad w_f(u) = -\frac{f'\big[F^{-1}(u)\big]}{f\big[F^{-1}(u)\big]}. \end{aligned} \tag{7}$$

Dabei benötigen wir die folgenden Voraussetzungen

$f(x) > 0$ für alle $x \in \mathbb{R}$ ,

Dichte $f$ und Gewichtsfunktion $w$ differenzierbar,

$w\big[F(x)\big]\, f(x) \to 0$ für $x \to \pm\infty$ .

Somit gilt beim Grenzübergang (1)

$$\frac{\mu - \mu_0}{\sigma_0} \longrightarrow \frac{\delta}{\tau\sqrt{\lambda(1-\lambda)}} \times \lambda(1-\lambda)\, I_f = \delta \sqrt{\lambda(1-\lambda)}\, \frac{I_f}{\tau}\,.$$

Wir erhalten also als Grenzwert der Gütefunktion

$$\begin{aligned} \beta_N(\delta) = P\left\{\frac{\bar{S}_Y - \mu}{\sigma} > \frac{k_r - \mu}{\sigma}\right\} \longrightarrow \beta(\delta) &= P\left\{N(0,1) > z_{1-\alpha} - c\,\delta\sqrt{\lambda(1-\lambda)}\right\} \\ &= P\left\{N(0,1) < -z_{1-\alpha} + c\,\delta\sqrt{\lambda(1-\lambda)}\right\}, \end{aligned} \tag{8}$$

wobei

$$c = \frac{I_f}{\tau} = \frac{\int_0^1 w(u)\, w_f(u)\, du}{\left(\int_0^1 \big(w(u) - \tilde{w}\big)^2\, du\right)^{\frac{1}{2}}} \quad \text{mit} \quad w_f(u) = -\frac{f'\big[F^{-1}(u)\big]}{f\big[F^{-1}(u)\big]}. \tag{9}$$

Beachte, dass die Konstante $c$ unverändert bleibt bei einer Lineartransformation der Gewichte $w(u) \to a\, w(u) + b$ , denn $\int w_f(u)\, du = 0$ (Substitution: $u = F(x)$ ). Weiter bleibt $c$ auch unverändert bei einer Lageänderung $f(x) \to f(x+b)$ ; bei einer Skalenänderung $f(x) \to f(x/a)/a$ gilt jedoch $w_f(u) \to w_f(u)/a$ , und damit ergibt sich $c \to c/a$ . Beim doppelten t-Test ändert sich die Konstante $c$ auf die gleiche Weise, und damit bleibt die relative Effizienz beim Vergleich mit einem linearen Rangtest invariant bei Lage- und Skalenänderungen.

*Frage:*
Wie ist bei gegebenen Dichte $f(x)$ die Gewichtsfunktion $w(u)$ zu wählen, damit der lineare Rangtest möglichst trennscharf wird?

*Antwort:*
Die Gewichtsfunktion ist so zu wählen, dass die Konstante $c$ maximal wird. Dies ist der Fall, falls $w(u) = w_f(u)$ . Es ist dann $\tilde{w} = 0$ und $c = \left(\int_0^1 w_f^2(u)\, du\right)^{\frac{1}{2}}$ . Bei gegebener Dichtefunktion $f$ besitzt also der lineare Rangtest mit der Gewichtsfunktion $w_f(u)$ unter allen linearen Rangtests die bestmögliche asymptotische Gütefunktion (Beweis im Anhang A4).

## 6. Asymptotische relative Effizienz, Beispiele

In diesem Abschnitt wollen wir die asymptotische relative Effizienz von einigen linearen Rangtests im Vergleich zum doppelten t-Test untersuchen. Wir haben sowohl beim doppelten t-Test als auch bei den linearen Rangtests den gleichen Grenzübergang (1) durchgeführt und dabei jeweils Grenzfunktionen von der folgenden Form gefunden:

$$\beta_N(\delta) \longrightarrow P\left\{N(0,1) < -z_{1-\alpha} + c\,\delta\sqrt{\lambda(1-\lambda)}\right\},$$

wobei die Konstante *c* abhängig ist vom jeweiligen Test. Beim t-Test ist gemäß (3) $c = c_t = 1/\sigma$, und bei den linearen Rangtests ist die Konstante *c* durch (9) gegeben, wobei das Integral $I_f$ gemäß (7) auf unterschiedliche Weise berechnet werden kann.

a) *Test von Wilcoxon*

Hier ist

$$w(u) = u - \tfrac{1}{2};$$

$$w_i = w\left(\frac{i}{N+1}\right) = \frac{i}{N+1} - \tfrac{1}{2} \quad \text{für } i = 1,\ldots,N \qquad (\text{äquivalent zu } w_i = i);$$

$$\int_0^1 w(u)\,du = 0, \quad \tau^2 = \int_0^1 w^2(u)\,du = \tfrac{1}{12};$$

$$c = c_W = I_f/\tau = \sqrt{12}\int_{-\infty}^{\infty} w'(F(x))\,f^2(x)\,dx = \sqrt{12}\int_{-\infty}^{\infty} f^2(x)\,dx.$$

Die asymptotische relative Effizienz des Tests von Wilcoxon im Vergleich zum doppelten t-Test beträgt also

$$are(W:t) = \left(\frac{c_W}{c_t}\right)^2 = 12\sigma^2\left(\int f^2(x)\,dx\right)^2,$$

wobei $\sigma^2 = Var(X_i) = Var(Y_j)$. Es gilt $are(W:t) \geq 108/125 = 0.864$ (Beweis im Anhang A5). Die asymptotische relative Effizienz des Tests von Wilcoxon im Vergleich zum t-Test kann beliebig gross werden ($\sigma^2 \to \infty$), sie kann aber nie schlechter werden als 0.864!

Für die logistische Verteilung ist

$$F(x) = \frac{1}{1+e^{-x}}, \quad f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\frac{f'(x)}{f(x)} = \frac{1-e^{-x}}{1+e^{-x}} = 2F(x) - 1;$$

es gilt

$$x = F^{-1}(u) \;\Leftrightarrow\; F(x) = u \;\Leftrightarrow\; e^{-x} = \frac{1}{u} - 1,$$

und wir erhalten

$$w_f(u) = -\frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)} = 2u - 1.$$

Diese Gewichtsfunktion ist äquivalent zur Gewichtsfunktion des Tests von Wilcoxon, und somit ist der Test von Wilcoxon der optimale lineare Rangtest bei der logistischen Verteilung.

b) *Normal-Scores-Test (Test mit Normal-Gewichten)*

Hier ist

$w(u)=\Phi^{-1}(u)$, wobei $\Phi=$ Verteilungsfunktion zu $N(0,1)$;

$w_i=w\left(\frac{i}{N+1}\right)=\Phi^{-1}\left(\frac{i}{N+1}\right)$ für $i=1,\ldots,N$;

$$\int_0^1 w(u)\,du=\int_0^1 \Phi^{-1}(u)\,du=\int_{-\infty}^{\infty} z\,\varphi(z)\,dz=0 \qquad \text{(Substitution: } u=\Phi(z)\text{)};$$

$$\tau^2=\int_0^1 w^2(u)\,du=\int_0^1 \left(\Phi^{-1}(u)\right)^2 du=\int_{-\infty}^{\infty} z^2\varphi(z)\,dz=1 \quad \text{(Substitution: } u=\Phi(z)\text{)};$$

$$c=c_{NS}=\frac{I_f}{\tau}=I_f=\int_0^1 w(u)\,w_f(u)\,du=\int_0^1 \Phi^{-1}(u)\,w_f(u)\,du.$$

Die asymptotische relative Effizienz des Normal-Scores-Tests im Vergleich zum t-Test beträgt

$$are(NS:t)=\left(\frac{c_{NS}}{c_t}\right)^2=\sigma^2 c_{NS}^2=\sigma^2\left(\int_0^1 \Phi^{-1}(u)\,w_f(u)\,du\right)^2.$$

Im Spezialfall der Normalverteilung ist $f=\varphi$, und da $\varphi'(z)/\varphi(z)=-z$, so gilt dann $w_f(u)=\Phi^{-1}(u)$; wir erhalten dann

$$c_{NS}=\int_0^1 \Phi^{-1}(u)\,w_f(u)\,du=\int_0^1 \left(\Phi^{-1}(u)\right)^2 du=\int_{-\infty}^{\infty} z^2\varphi(z)\,dz=1,$$

und damit ist im Falle einer Normalverteilung $are(NS:t)=1$. Bei Vorliegen einer Normalverteilung ist also der Normal-Scores-Test asymptotisch äquivalent zum doppelten t-Test.

Es gilt $are(NS:t)\geq 1$ (Beweis im AnhangA6). Die asymptotische relative Effizienz des Normal-Scores-Tests im Vergleich zum t-Test kann also nie schlechter werden als 1! Dabei gilt das Gleichheitszeichen nur im Falle der Normalverteilung; bei jeder anderen Verteilung ist der Normal-Scores-Test (asymptotisch) besser als der t-Test.

Im Spezialfall der Normalverteilung ist $f=\varphi$, und $w_f(u)=\Phi^{-1}(u)$, wie wir oben gesehen haben; die Gewichte des Normal-Scores-Tests sind also optimal im Falle einer Normalverteilung.

c) *Vorzeichentest von Mood (Mediantest von Mood)*

Hier ist

$$w(u)=\begin{cases}-1 & \text{für } u<\frac{1}{2}\\ 0 & \text{für } u=\frac{1}{2}\\ +1 & \text{für } u>\frac{1}{2};\end{cases}$$

falls $N=2k$, d.h. falls $N$ gerade: $w_i=w\left(\frac{i}{N+1}\right)=\begin{cases}-1 & \text{für } i\leq k\\ +1 & \text{für } i>k;\end{cases}$

falls $N=2k+1$, d.h. falls $N$ ungerade: $w_i=w\left(\frac{i}{N+1}\right)=\begin{cases}-1 & \text{für } i\leq k\\ 0 & \text{für } i=k+1\\ +1 & \text{für } i>k+1.\end{cases}$

Es sei

$Ze =$ Zentralwert der Gesamtstichprobe $X_1,\ldots,X_m,Y_1,\ldots Y_n$;

$N_+ = \#\left\{ j \,\middle|\, Y_j > Ze \right\}$ = Anzahl der $Y$-Werte mit Gewicht +1;

$N_- = \#\left\{ j \,\middle|\, Y_j < Ze \right\}$ = Anzahl der $Y$-Werte mit Gewicht $-1$.

Dann gilt

$S_Y = \sum_{j=1}^{n} w_{R_j} = \sum_{j=1}^{n} gew(Y_j) = N_+ - N_-$ (ähnlich wie beim Vorzeichentest).

Es gilt

$$\int_0^1 w(u)\,du = 0, \quad \tau^2 = \int_0^1 w^2(u)\,du = 1.$$

Nun ist hier die Gewichtsfunktion $w(u)$ unstetig bei $u = \frac{1}{2}$; daher approximieren wir die Funktion $w$ mit einer differenzierbaren Funktion $\tilde{w}$, für welche $\tilde{w}'(u) = 0$ für $\left|u - \frac{1}{2}\right| > \varepsilon$, und lassen $\varepsilon$ gegen Null streben. Wir finden so

$$c = c_V = \frac{I_f}{\tau} = I_f = \lim_{\varepsilon \to 0} \int_0^1 \tilde{w}'(u) f\left(F^{-1}(u)\right) du = f\left(F^{-1}(\tfrac{1}{2})\right) \lim_{\varepsilon \to 0} \int_0^1 \tilde{w}'(u)\,du = 2 f(\mu),$$

wobei $\mu = F^{-1}(\frac{1}{2}) =$ Median von $F$. Die asymptotische relative Effizienz des Vorzeichentests von Mood im Vergleich zum t-Test beträgt somit

$$are(V:t) = \left(\frac{c_V}{c_t}\right)^2 = 4\sigma^2 f^2(\mu) \quad (\mu = \text{ Median von } f).$$

Es gilt $0 \le are(V:t) \le \infty$. Wir haben hier das gleiche Resultat wie beim Vergleich des einfachen t-Tests mit den einfachen Vorzeichentest (vgl. Kapitel 2).

Die zweiseitige Exponentialverteilung hat die Dichte $f(x) = \frac{1}{2} e^{-|x|}$, und es gilt

$$-\frac{f'(x)}{f(x)} = \begin{cases} +1 & \text{für } x > 0 \\ -1 & \text{für } x < 0, \end{cases}$$

und somit gilt

$$w_f(u) = -\frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)} = \begin{cases} +1 & \text{für } u > \frac{1}{2} \\ -1 & \text{für } u < \frac{1}{2}. \end{cases}$$

Dies zeigt, dass die Gewichtung des Vorzeichentests von Mood optimal ist bei der zweiseitigen Exponentialverteilung.

d) *Optimale Gewichtung bei einer Cauchy-Verteilung*; *Cauchy-Scores-Test*

Die Dichtefunktion der Cauchy-Verteilung ist gegeben durch

$$f(x)=\frac{1}{\pi(1+x^2)},\ -\infty<x<\infty.$$

Es gilt

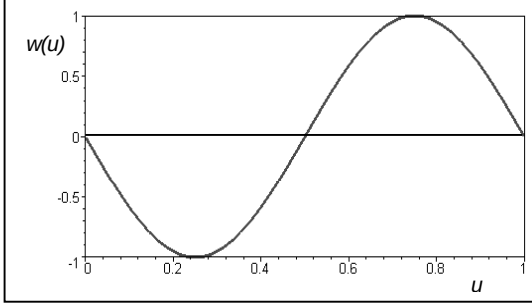
$$-\frac{f'(x)}{f(x)}=\frac{2x}{1+x^2}$$

$$F(x)=\frac{1}{\pi}\left(\arctan(x)+\tfrac{1}{2}\right)$$

$$F^{-1}(u)=-\operatorname{cotan}(\pi u)$$

$$w_f(u)=-\frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}=-\sin(2\pi u).$$

Abb. 1: Gewichtsfunktion $w(u)=-\sin(2\pi u)$




Die optimale Gewichtsfunktion ist also gegeben durch $w(u)=-\sin(2\pi u)$ (vgl. Abb. 1). Extreme Beobachtungen erhalten geringes Gewicht, während Beobachtungen in der Umgebung des ersten und dritten Quartils großes negatives bzw. positives Gewicht erhalten.

Wir betrachten nun den linearen Rangtest mit der Gewichtsfunktion $w(u)=-\sin(2\pi u)$; die zugehörigen Gewichte nennen wir Cauchy-Gewichte (Cauchy-Scores), weil sie optimal sind bei einer Cauchy-Verteilung:

$$w(u)=-\sin(2\pi u);$$

$$w_i=w\left(\frac{i}{N+1}\right)=-\sin\left(\frac{2\pi i}{N+1}\right)\ \text{für}\ i=1,\ldots,N;$$

$$\int_0^1 w(u)\,du=-\int_0^1 \sin(2\pi u)\,du=0;$$

$$\tau^2=\int_0^1 w^2(u)\,du=\int_0^1 \sin^2(2\pi u)\,du=\tfrac{1}{2};$$

$$c=c_{CS}=\frac{I_f}{\tau}=\sqrt{2}I_f=\sqrt{2}\int_0^1 w(u)\,w_f(u)\,du=-\sqrt{2}\int_0^1 \sin(2\pi u)\,w_f(u)\,du.$$

Die asymptotische relative Effizienz des Cauchy-Scores-Tests im Vergleich zum t-Test beträgt

$$are(CS:t)=\left(\frac{c_{CS}}{c_t}\right)^2=\sigma^2c_{CS}^2=2\sigma^2\left(\int_0^1 \sin(2\pi u)\,w_f(u)\,du\right)^2.$$

*Vergleichende Übersicht über unsere Beispiele*

Wir betrachten die folgenden vier Verteilungen für unser Lagemodell:

a) Normalverteilung $N(0,1)$: $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2},\ -\infty<x<\infty$

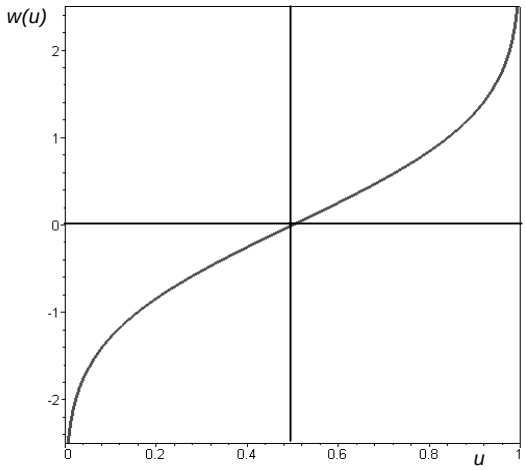
b) Logistische Verteilung: $f(x)=\frac{e^{-x}}{\left(1+e^{-x}\right)^2},\ -\infty<x<\infty$

c) Laplaceverteilung: $f(x)=\frac{1}{2}e^{-|x|},\ -\infty<x<\infty$
(zweiseitige Exponentialverteilung)

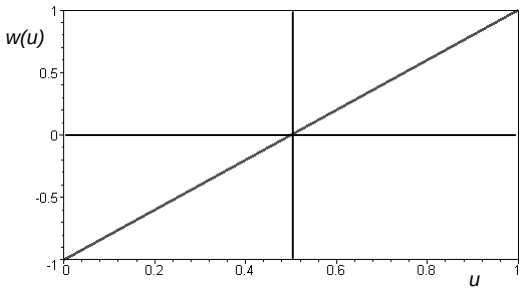
d) Cauchyverteilung $C(0,1)$: $f(x)=\frac{1}{\pi(1+x^2)},\ -\infty<x<\infty$

Weiter betrachten wir den klassischen t-Test und die folgenden vier linearen Rangtests:

1) Normal-Scores-Test
mit Gewichtsfunktion $w(u)=\Phi^{-1}(u)$

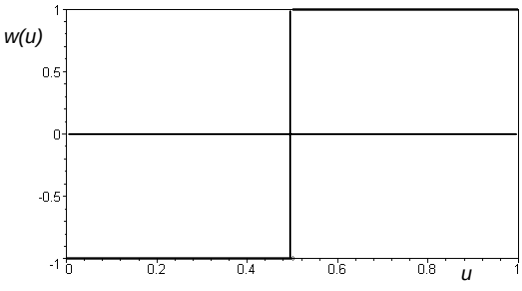



2) Test von Wilcoxon
mit Gewichtsfunktion $w(u)=2u-1$




3) Vorzeichentest von Mood
mit Gewichtsfunktion

$$w(u)=\begin{cases}-1 & \text{für } u<1/2\\ 0 & \text{für } u=1/2\\ +1 & \text{für } u>1/2\end{cases}$$




4) Cauchy-Scores-Test
mit Gewichtsfunktion $w(u)=-\sin(2\pi u)$

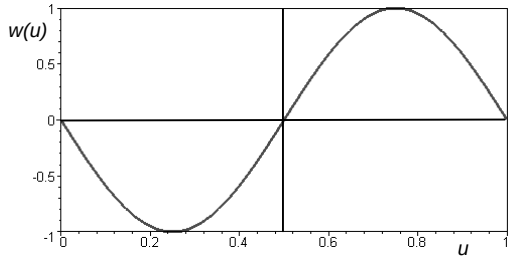


**Tabelle 1:** *Konstante c bei asymptotischer Gütefunktion:*

$$\beta_N(\delta) \to P\left\{N(0,1) < -z_{1-\alpha} + c\,\delta\sqrt{\lambda(1-\lambda)}\right\}.$$

t-Test: $c = c_t$, Vorzeichentest: $c = c_V$, Wilcoxontest: $c = c_W$,
Normal-Scores-Test: $c = c_{NS}$, Cauchy-Scores-Test: $c = c_{CS}$

| | $c_t^2$ | $c_{NS}^2$ | $c_W^2$ | $c_V^2$ | $c_{CS}^2$ | Rangfolge |
|---|---|---|---|---|---|---|
| Normal-verteilung | 1 | 1 | $\frac{3}{\pi} = 0.955$ | $\frac{2}{\pi} = 0.637$ | 0.430 | $c_t = c_{NS} > c_W > c_V > c_{CS}$ |
| Logistische Verteilung | $\frac{3}{\pi^2} = 0.304$ | 0.318 | $\frac{1}{3}$ | $\frac{1}{4}$ | 0.203 | $c_W > c_{NS} > c_t > c_V > c_{CS}$ |
| Laplace-verteilung | $\frac{1}{2}$ | $\frac{2}{\pi} = 0.637$ | $\frac{3}{4}$ | 1 | $\frac{8}{\pi^2} = 0.811$ | $c_V > c_{CS} > c_W > c_{NS} > c_t$ |
| Cauchy-verteilung | 0 | 0.215 | $\frac{3}{\pi^2} = 0.304$ | $\frac{4}{\pi^2} = 0.405$ | $\frac{1}{2}$ | $c_{CS} > c_V > c_W > c_{NS} > c_t$ |

*Maple-Arbeitsblatt* `asreleff.mws` *zur Berechnung der Werte in Tabelle 1*

```
> #
  # Asymptotische relative Effizienz zwischen
  #   t-Test, Vorzeichentest, Wilcoxontest,
  #   Normal-Scores-Test und Cauchy-Scores-Test
 #
> restart;
> Phi := z -> 1/2 + (1/2)*erf(z/sqrt(2)); # Verteilungsfunktion zu N(0,1)
> Phiinv := u -> solve(Phi(z)=u,z);       # Inverse Verteilungsfunktion zu N(0,1)
> # Dichtefunktionen der 4 Verteilungen
> #f := x -> exp(-x^2/2)/sqrt(2*Pi);  # Normalverteilung
> #f := x -> exp(-x)/(1+exp(-x))^2;   # Logistische Verteilung
> #f := x -> (1/2)*exp(-abs(x));      # Laplaceverteilung
> #f := x -> 1/(Pi*(1+x^2));          # Cauchyverteilung
> plot(f(x),x=-3..3);
> int(f(x),x=-infinity..infinity);
> # Optimale Gewichtsfunktion wf(u) bei den 4 Verteilungen
> #wf := u -> Phiinv(u);                # Normalverteilung
> #wf := u -> 2*u-1;                    # Logistische Verteilung
> #wf := u -> abs(1,2*u-1);             # Laplaceverteilung
> #wf := u -> -sin(2*Pi*u);             # Cauchyverteilung
> plot(wf(u),u=0..1,view=[0..1,-3..3]);
> # c_t = Konstante beim t-Test
> sigma2 := int(x^2*f(x),x=-infinity..infinity);
> c_t := 1/sqrt(sigma2);
> c_t^2;
> evalf(%);
> # c_V = Konstante beim Vorzeichentest
> c_V := (2*f(0));
> c_V^2;
> evalf(%);
> # c_W = Konstante beim Test von Wilcoxon
> c_W := 2*sqrt(3)*int(f(x)^2,x=-infinity..infinity);
> c_W^2;
> evalf(%);
> # c_NS = Konstante bei Normal-Score-Test
> # Gewichtsfunktion beim Normal-Score-Test
  w := u -> Phiinv(u);
> #plot(w(u),u=0..1,view=[0..1,-3..3]);
> c_NS := evalf(Int(w(u)*wf(u),u=0.0001..0.9999));
> c_NS := evalf(2*Int(w(u)*wf(u),u=0.00001..1/2));
> #c_NS :=  Re(%);
> c_NS^2;
> # c_CS = Konstante beim Cauchy-Scores-Test
> w := u -> -sin(2*Pi*u);
> c_CS := evalf(sqrt(2)*Int(w(u)*wf(u),u=0..1));
```

```
  c_CS := evalf(2*sqrt(2)*Int(w(u)*wf(u),u=1/2..1));
> #c_CS := Re(%);
> c_CS^2;
> are(t.V) := evalf((c_t/c_V)^2);
> are(t.W) := evalf((c_t/c_W)^2);
> are(t.NS) := evalf((c_t/c_NS)^2);
> are(t.CS) := evalf((c_t/c_CS)^2);
> are(NS.W) := evalf((c_NS/c_W)^2);
> are(NS.V) := evalf((c_NS/c_V)^2);
> are(NS.CS) := evalf((c_NS/c_CS)^2);
> are(W.V) := (c_W/c_V)^2;
> are(W.CS) := evalf((c_W/c_CS)^2);
> are(V.CS) := evalf((c_V/c_CS)^2);
> ###
> # c_NS bei Laplace-Verteilung
> phi := z -> exp(-z^2/2)/sqrt(2*Pi);
> c_NS := 2*int(z*phi(z),z=0..infinity);
> c_NS^2;
> evalf(%);
> # c_CS bei Laplace-Verteilung
> c_CS := 2*int(sin(2*Pi*u),u=0..1/2);
> c_CS^2;
> evalf(%);
>
```

## 7. Lokal optimale Rangtests (locally most powerful rank tests)

Lit. Hettmansperger (1984), S. 142-147.

Wir haben bisher optimale lineare Rangtests aufgrund der asymptotischen Gütefunktion bestimmt. Ein anderer Weg optimiert die Gütefunktion von linearen Rangtests in der Umgebung von $\vartheta = 0$ bei endlichen Stichprobenumfängen, und dies führt zu sogenannten *lokal optimalen Rangtests (locally most powerful rank tests)*. Die optimale Gewichtung dieser Tests ist gegeben durch

$$(10)\quad a_i = -E\left(\frac{f'\left(V_{(i)}\right)}{f\left(V_{(i)}\right)}\right)$$

wobei $V_1,\ldots,V_N$ unabhängige Zufallsgrößen je mit der Dichte *f* und $V_{(1)},\ldots,V_{(N)}$ die zugehörigen Ordnungsstatistiken sind. Man beachte die enge Verwandtschaft mit der optimalen asymptotischen Gewichtung, die wir im Abschnitt 6 gefunden haben:

$$(11)\quad w_i = -\frac{f'\left(F^{-1}\left(\frac{i}{N+1}\right)\right)}{f\left(F^{-1}\left(\frac{i}{N+1}\right)\right)}.$$

Wir wollen plausibel machen, dass die beiden Gewichtungen asymptotisch äquivalent sind. Es seien $U_1,\ldots,U_N$ unabhängige Zufallsgrößen je mit einer Gleichverteilung $U(0,1)$, und $U_{(1)},\ldots,U_{(N)}$ seien die zugehörigen Ordnungsstatistiken. Wir setzen $V_i = F^{-1}(U_i)$ für $i = 1,\ldots,N$; dann sind $V_1,\ldots,V_N$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion *F* und der Dichte *f*, und die Ordnungsstatistiken $V_{(1)},\ldots,V_{(N)}$ sind gegeben durch $V_{(i)} = F^{-1}(U_{(i)})$, $i = 1,\ldots,N$. Da die Varianz der Ordnungsstatistiken $V_{(i)}$ für $N \to \infty$ gegen Null konvergiert (zumindest falls *i* nicht extrem groß oder extrem klein ist), so gilt $E\left(h(V_{(i)})\right) \to h\left(E(V_{(i)})\right)$ für jede differenzierbare Funktion *h*, und damit gilt

$$a_i = -E\left(\frac{f'\left(V_{(i)}\right)}{f\left(V_{(i)}\right)}\right) \longrightarrow -\frac{f'\left(E(V_{(i)})\right)}{f\left(E(V_{(i)})\right)}.$$

Nun gilt weiter

$$E(V_{(i)}) = E\left(F^{-1}(U_{(i)})\right) \longrightarrow F^{-1}\left(E(U_{(i)})\right) = F^{-1}\left(\frac{i}{N+1}\right),$$

denn es gilt $E(U_{(i)}) = i/(N+1)$ (vgl. Anhang A7). Somit gilt $|a_i - w_i| \to 0$ für $N \to \infty$, falls *i* nicht extrem groß oder extrem klein ist; damit wird plausibel, dass die beiden Tests asymptotisch äquivalent sind.

*Beispiel 1: Normalverteilung*

Falls unsere Zufallsgrößen normaverteilt sind, so ist $f = \varphi$ und $-f'(x)/f(x) = x$, und wir erhalten

$$a_i = -E\left(Z_{(i)}\right),$$

wobei $Z_1,\ldots,Z_N$ unabhängige Zufallsgrößen je mit einer standardisierten Normalverteilung und $Z_{(1)},\ldots,Z_{(N)}$ die zugehörigen Ordnungsstatistiken sind. Offensichtlich sind die Gewichte $a_i$ und $w_i$ antisymmetrisch verteilt: $a_i = -a_{N-i+1}$, $w_i = -w_{N-i+1}$ für $i = 1,\ldots,N$ (vgl. Tabelle 2). Man nennt den Test mit den Gewichten $a_i$ den *Normal-Scores-Test nach Fisher-Yates*, während der Test mit den Gewichten $w_i$ als Normal-Scores-Test nach van der Waerden (X-Test) bezeichnet wird. Die beiden Tests sind asymptotisch äquivalent. Ob eine der beiden Testversionen bei endlichen Stichproben der anderen eindeutig überlegen ist, ist für mich eine noch offene Frage.

**Tabelle 2:** Gewichte $a_i$ und $w_i$ bei der Normalverteilung

a) für $N = 10$

| $i$ | $w_i$ | $a_i$ |
|---|---|---|
| 1 | –1.335 | –1.539 |
| 2 | –0.908 | –1.001 |
| 3 | –0.605 | –0.656 |
| 4 | –0.349 | –0.376 |
| 5 | –0.114 | –0.123 |
| 6 | 0.114 | 0.123 |
| 7 | 0.349 | 0.376 |
| 8 | 0.605 | 0.656 |
| 9 | 0.908 | 1.001 |
| 10 | 1.335 | 1.539 |

b) für $N = 20$

| $i$ | $w_i$ | $a_i$ |
|---|---|---|
| 1 | –1.668 | –1.867 |
| 2 | –1.309 | –1.408 |
| 3 | –1.068 | –1.131 |
| 4 | –0.876 | –0.921 |
| 5 | –0.712 | –0.745 |
| 6 | –0.566 | –0.590 |
| 7 | –0.431 | –0.448 |
| 8 | –0.303 | –0.315 |
| 9 | –0.180 | –0.187 |
| 10 | –0.060 | –0.062 |
| 11 | 0.060 | 0.062 |
| ⋮ | ⋮ | ⋮ |

*Beispiel 2: Logistische Verteilung*

Falls unsere Zufallsgrößen eine logistische Verteilung besitzen, so gilt $-f'(x)/f(x) = 2F(x) - 1$ (vgl. oben), und da $F(V_i) = U_i$ eine Gleichverteilung $U(0,1)$ besitzt, so finden wir

$$a_i = -E\left(\frac{f'\left(V_{(i)}\right)}{f\left(V_{(i)}\right)}\right) = E\left(2F(V_{(i)}) - 1\right) = 2E\left(U_{(i)}\right) - 1,$$

wobei $U_1,\ldots,U_N$ unabhängige Zufallsgrößen je mit einer Gleichverteilung $U(0,1)$ und $U_{(1)},\ldots,U_{(N)}$ die zugehörigen Ordnungsstatistiken sind. Nun gilt $E(U_{(i)}) = i/(N+1)$ (vgl. Anhang A7), und daher ist

$$a_i = 2\frac{i}{N+1} - 1 = w_i;$$

die beiden Gewichtungen sind also identisch, und somit ist der Test von Wilcoxon bei der logistischen Verteilung nicht nur asymptotisch optimal sondern auch lokal optimal.

*Beispiel 3: Laplaceverteilung (zweiseitige Exponentialverteilung)*

Falls unsere Zufallsgrößen eine Laplaceverteilung besitzen, so gilt

$$-f'(x)/f(x)=\begin{cases}-1 & \text{für } x<0\\ +1 & \text{für } x>0\end{cases}$$

(für $x=0$ ist die Ableitung nicht definiert), und wir erhalten $a_i = E\left(\text{sgn}(V_{(i)})\right)$, wobei $\text{sgn}(x)$ das Vorzeichen von $x$ bezeichnet; der Wert von dieser Funktion für $x=0$ ist hier irrelevant, da die Zufallsgröße $V_{(i)}$ stetig ist. Nun sei $Y=\#\left\{j\middle|V_j>0\right\}$; dann ist $Y \sim Bi(N,\frac{1}{2})$, und es gilt $\left\{V_{(i)}>0\right\}=\left\{Y>i\right\}$. Somit haben wir

$$a_i = E\left(\text{sgn}(V_{(i)})\right)=P\left\{V_{(i)}>0\right\}-P\left\{V_{(i)}<0\right\}=2P\left\{V_{(i)}>0\right\}-1=2P\left\{Bi(N,\tfrac{1}{2})>i\right\}-1.$$

Tabelle 3 zeigt die Gewichte $a_i$ und $w_i$ für $N=10$ und $N=20$. Die Frage, ob die Gewichtung $a_1,\ldots,a_N$ bei endlichen Stichprobenumfängen der asymptotisch optimalen Gewichtung $w_1,\ldots,w_N$ eindeutig überlegen ist, ist meines Wissens noch nicht geklärt.

**Tabelle 3:** Gewichte $a_i$ und $w_i$ bei der Laplace-Verteilung

a) für $N=10$

| $i$ | $w_i$ | $a_i$ |
|---|---|---|
| 1 | –1 | –0.998 |
| 2 | –1 | –0.979 |
| 3 | –1 | –0.891 |
| 4 | –1 | –0.656 |
| 5 | –1 | –0.246 |
| 6 | 1 | 0.246 |
| 7 | 1 | 0.656 |
| 8 | 1 | 0.891 |
| 9 | 1 | 0.979 |
| 10 | 1 | 0.998 |

b) für $N=20$

| $i$ | $w_i$ | $a_i$ |
|---|---|---|
| 1 | –1 | –1.000 |
| 2 | –1 | –1.000 |
| 3 | –1 | –1.000 |
| 4 | –1 | –0.997 |
| 5 | –1 | –0.988 |
| 6 | –1 | –0.959 |
| 7 | –1 | –0.885 |
| 8 | –1 | –0.737 |
| 9 | –1 | –0.497 |
| 10 | –1 | –0.176 |
| 11 | 1 | 0.176 |
| ⋮ | ⋮ | ⋮ |

*Maple-Arbeitsblatt* `norscores.mws` *zur Berechnung der Werte in Tabelle 2*

```
> #
  # Normal-Scores nach Methode 1 und 2
  #
> restart;
> Phi:=z->1/2+1/2*erf(z/sqrt(2));
> Phic:=z->(1/2)*erfc(z/sqrt(2));
> phi:=z->exp(-z^2/2)/sqrt(2*Pi);
> #
  # Normal Scores:
  # 1. w(i) = Phiinv(i/(N+1))
  # 2. a(i) = E(Z_(i)), wobei Z1,...,ZN unabhängig, je mit N(0,1)
  #
> N := 10;
> # 1. Methode:
> w := i -> fsolve(Phi(z)=i/(N+1),z);
> w(3);
> # 2. Methode:
  # F = Verteilungsfunktion von Z_(i)
  # F(i,x) = P{Z_(i) < x} = P{Nx >= i}, wobei Nx = #{i|Zi < x}
  # Nx ~ Bi(N,p), p = Phi(x)
> F := (i,x) -> sum(binomial(N,j)*Phi(x)^j*Phic(x)^(N-j),j=i..N);
> plot(F(3,x),x=-3..3);
> f := (i,x) -> diff(F(i,x),x);
> a := i -> Int(x*f(i,x),x=-infinity..infinity);
> a(3);
> evalf(%);
> N:=20;
> for i from 1 to N do
    print(i,w(i),evalf(a(i)));
  end do:
>
```

## 8. Permutationstest

Wir betrachten wiederum unser Zwei-Stichproben-Problem mit den beiden Stichproben $X_1,\ldots,X_m$ und $Y_1,\ldots,Y_n$. Nun wird die Gesamtstichprobe nach Größe geordnet zu $V_1,\ldots,V_N$ mit $N=m+n$, und die Realisationen $v_1,\ldots,v_N$ werden als Gewichte eines linearen Rangtests benutzt; der lineare Rangtest mit diesen natürlichen Gewichten heißt *Permutationstest*. Die Gewichte des Permutationstests berücksichtigen in natürlicher Weise die unbekannte Verteilung unserer Zufallsgrößen, und man könnte hoffen, dass der Test deshalb ein asymptotisch optimaler linearer Rangtest ist bei beliebiger Verteilungsfunktion *F*. Dies ist jedoch nicht der Fall, der Test ist asymptotisch äquivalent zum klassischen doppelten t-Test, und er ist daher wie der t-Test nur optimal bei Vorliegen einer Normalverteilung.

*Begründung:*
Die Zufallsgrößen $X_1,\ldots,X_m$ besitzen je die Dichte $f(x)$ und die Zufallsgrößen $Y_1,\ldots,Y_n$ je die Dichte $f(x-\vartheta)$. Unter $H_0\colon \vartheta=0$ besitzen alle Zufallsgrößen die gleiche Dichte *f* mit der Verteilungsfunktion *F*. Wir betrachten drei lineare Rangtests mit den folgenden Gewichtungen:

A) $w_i=v_i$ für $i=1,\ldots,n$ (Permutationstest);

B) $w_i=E(V_i)$ für $i=1,\ldots,n$;

C) $w_i=F^{-1}\left(\dfrac{i}{N+1}\right)$ für $i=1,\ldots,n$.

Es ist plausibel, dass Test A und B asymptotisch äquivalent sind, und wir haben uns im Abschnitt 7 überlegt, dass Test B asymptotisch äquivalent ist zu Test C; somit ist Test A äquivalent zu Test C. Wir berechnen nun die Konstante $c$ der asymptotischen Gütefunktion des Tests C. Gemäß (9) gilt

$$c=\frac{I_f}{\tau}=\frac{\int_0^1 w(u)\,w_f(u)\,du}{\left(\int_0^1\left(w(u)-\tilde{w}\right)^2 du\right)^{\frac{1}{2}}} \quad \text{mit} \quad w_f(u)=-\frac{f'\left[F^{-1}(u)\right]}{f\left[F^{-1}(u)\right]}.$$

Beim Test C ist die Gewichtsfunktion gegeben durch $w(u)=F^{-1}(u)$, und wir erhalten

$$\tilde{w}=\int_0^1 w(u)\,du=\int_0^1 F^{-1}(u)\,du=\int_{-\infty}^{\infty} x\,f(x)\,dx=\mu=E(X_i),$$

$$\tau^2=\int_0^1\left(w(u)-\mu\right)^2 du=\int_{-\infty}^{\infty}\left(x-\mu\right)^2 f(x)\,dx=\sigma^2=Var(X_i),$$

$$\int_0^1 w(u)\,w_f(u)\,du=-\int_0^1 F^{-1}(u)\frac{f'\left[F^{-1}(u)\right]}{f\left[F^{-1}(u)\right]}du=-\int_{-\infty}^{\infty} x\,f'(x)\,dx=\int_{-\infty}^{\infty} f(x)\,dx=1,$$

und damit ist die Konstante *c* unserer drei Tests gegeben durch $c=1/\sigma$ wie beim klassischen t-Test (vgl. Abschnitt 4). Der Permutationstest ist also asymptotisch äquivalent zum klassischen doppelten t-Test. Ein Vorteil im Vergleich zum t-Test besteht darin, dass der Permuationstest das vorgegebene Signifikanzniveau auch einhält, wenn keine Normalverteilung vorliegt, die asymptotische Optimalität ist jedoch nur bei einer Normalverteilung gegeben.

# 8. Test von Wilcoxon-Mann-Whithey bei beliebigen Verteilungen

## 1. Problemstellung, stochastische Ordnung

Wir betrachten hier das Zweistichprobenproblem in allgemeiner Form:

$\left.\begin{matrix} X_1,\ldots,X_m \\ Y_1,\ldots,Y_n \end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases} \text{je mit Verteilungsfunktion } F \\ \text{je mit Verteilungsfunktion } G. \end{cases}$

A) *Klassischer Fall*

$X_i \sim N(\mu_1,\sigma^2)$ und $Y_j \sim N(\mu_2,\sigma^2)$;

$H_0: \mu_1 = \mu_2$ (beide Stichproben haben gleiche Verteilung);

$H_1: \mu_1 < \mu_2$ (*Y*-Werte im Mittel größer als *X*-Werte);

$\rightarrow$ klassischer doppelter t-Test.

B) *Allgemeines Lagemodell*

$X_i \sim F$ und $Y_j \sim G$, wobei $G(x) = F(x-\vartheta)$, $Y_j - \vartheta$ hat gleiche Vert. wie $X_i$;

$H_0: \vartheta = 0$ (beide Stichproben haben gleiche Verteilung);

$H_1: \vartheta > 0$ (*Y*-Werte im Mittel größer als *X*-Werte).

C) *Stochastische Ordung (Typ I)*

$X_i \sim F$ und $Y_j \sim G$;

$H_0: F \equiv G$ (beide Stichproben haben gleiche Verteilung);

$H_1: F(x) \geq G(x)\ \forall x$ und $F \not\equiv G$ (*Y*-Werte im Mittel größer als *X*-Werte).

Beachte:

$$F(x) \geq G(x)\ \forall x \quad \Leftrightarrow \quad F^c(x) = 1 - F(x) \leq G^c(x)\ \forall x$$
$$\Leftrightarrow \quad P\{X > x\} \leq P\{Y > x\}\ \forall x, \text{ d.h. } Y\text{-Werte im Mittel } \textit{größer} \text{ als } X\text{-Werte.}$$

D) *Stochastische Ordung (Typ II)*

$X_i \sim F$ und $Y_j \sim G$;

$H_0: F \equiv G$ (beide Stichproben haben gleiche Verteilung);

$H_1: P\{X < Y\} > P\{X > Y\}$ (*Y*-Werte im Mittel größer als *X*-Werte).

Nun haben wir zwei Definitionen von stochastischer Ordnung:

*Definition I:*

*X* und *Y* seien zwei Zufallsgrößen mit den Verteilungsfunktionen *F* und *G*. Dann definieren wir die Ordnungsrelation $\overset{\text{I}}{\prec}$ *(stochastische Ordnung vom Typ I)* wie folgt:

$$X \overset{\text{I}}{\prec} Y \quad \Leftrightarrow \quad F(x) \geq G(x)\ \forall x \text{ und } F \not\equiv G.$$

*Definition II:*

*X* und *Y* seien zwei unabhängige Zufallsgrößen mit den Verteilungsfunktionen *F* und *G*. Dann definieren wir die Ordnungsrelation $\overset{\text{II}}{\prec}$ *(stochastische Ordnung vom Typ II)* wie folgt:

$$X \overset{\text{II}}{\prec} Y \quad \Leftrightarrow \quad P\{X < Y\} > P\{X > Y\}.$$

*Behauptung:*
Die stochastische Ordnung nach Typ I impliziert die Ordnung nach Typ II, d.h.

$$X \overset{\text{I}}{\prec} Y \;\Rightarrow\; X \overset{\text{II}}{\prec} Y .$$

*Beweis im stetigen Fall:*
$X$ und $Y$ seien unabhängige Zufallsgrößen mit den stetigen Verteilungsfunktionen $F$ und $G$. Dann gilt

(1) $P\{X<Y\} = \int F(x)\,dG(x), \quad P\{X=Y\} = 0, \quad P\{X>Y\} = \int G(x)\,dF(x).$

Falls beide Zufallsgrößen $X$ und $Y$ die gleiche Verteilungsfunktion $F$ besitzen, so gilt

$$P\{X<Y\} = P\{X>Y\} = \tfrac{1}{2}, \text{ und somit ist dann } \int F(x)\,dF(x) = \tfrac{1}{2}.$$

Nun sei $X \overset{\text{I}}{\prec} Y$, d.h. $F(x) \geq G(x)\ \forall x$ und $F \not\equiv G$. Dann folgt aus (1)

$$P\{X<Y\} = \int F(x)\,dG(x) > \int G(x)\,dG(x) = \tfrac{1}{2},$$

und somit ist

$$P\{X<Y\} > \tfrac{1}{2}, \text{ d.h. } P\{X<Y\} > P\{X>Y\};$$

damit ist $X \overset{\text{II}}{\prec} Y$, was zu beweisen war. Der Beweis für den allgemeinen Fall findet sich im Anhang A8. Die Umkehrung der obigen Behauptung ist jedoch nicht richtig, wie das folgende Gegenbeispiel zeigt.

*Beispiel:* Es sei $X$ eine Zufallsgröße mit einer Rechtecksverteiung in $[0,1]$, und $Y$ besitze die Dichte

$$g(y) = \begin{cases} \frac{1}{2} & \text{für } 0 \leq y \leq \frac{2}{5} \\ 2 & \text{für } \frac{2}{5} < y \leq \frac{4}{5} \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt $P\{X<Y\} = \int F(y)\,g(y)\,dy = \int y\,g(y)\,dy = \frac{13}{25} > \frac{1}{2}$, und somit ist $X \overset{\text{I}}{\prec} Y$. Es gilt jedoch nicht $X \overset{\text{II}}{\prec} Y$, da sich die beiden Verteilungsfunktionen $F$ und $G$ schneiden.

*Bemerkung zur Ordnung vom Typ II*
Aus $P\{X<Y\} = P\{X>Y\}$ folgt nicht $F \equiv G$. Gegenbeispiel: Es seien $X$ und $Y$ unabhängige Zufallsgrößen mit Gleichverteilungen im Intervall $[-1,1]$ bzw. $[-2,2]$. Dann gilt offensichtlich aus Symmetriegründen $P\{X<Y\} = P\{X>Y\}$, die beiden Verteilungsfunktionen sind jedoch verschieden.

Die Ordnung nach Typ II ist also feiner als die Ordnung nach Typ I. Das Testproblem gemäß D ist somit von den vier Varianten A bis D die allgemeinste:

A ist Spezialfall von B, B ist Spezialfall von C, C ist Spezialfall von D.

Wir werden uns im folgenden mit dem Testproblem D (allgemeinste Variante) befassen.

## 2. Test von Wilcoxon-Mann-Whitney im stetigen Fall

Wir wollen nochmals das Vorgehen im stetigen Fall skizzieren.

$\left.\begin{matrix} X_1,\ldots,X_m \\ Y_1,\ldots,Y_n \end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases} \text{je mit Verteilungsfunktion } F \\ \text{je mit Verteilungsfunktion } G. \end{cases}$

$H_0: F \equiv G$ (beide Stichproben haben gleiche Verteilung);

$H_1: P\{X < Y\} > P\{X > Y\}$ ($Y$-Werte im Mittel größer als $X$-Werte).

Wir betrachten die Testgröße

$$N_+ = \#\left\{(i,j)\middle| X_i < Y_j\right\}.$$

Die Nullhypothese wird abgelehnt, falls $N_+$ zu groß ist. Anstelle von $N_+$ kann man auch mit der Rangsumme arbeiten. Es sei

$R_j = rang(Y_j) =$ Rang von $Y_j$ in der geordneten Gesamtstichprobe;

$$S_Y = \sum_{j=1}^{n} R_j = \text{ Rangsumme der } Y\text{-Beobachtungen.}$$

Dann gilt

(2) $$S_Y = N_+ + \frac{n(n+1)}{2}.$$

Anstelle von $N_+$ (Testgröße von Mann) kann man also auch mit $S_Y$ (Testgröße von Wilcoxon) arbeiten. Wir werden im folgenden die Testgröße $N_+$ von Mann verwenden.

*Beweis von* (2):

Wir betrachten die Ordnungsstatistiken zur $Y$-Stichprobe: $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$. Dann gilt

$$rang(Y_{(1)}) = \#\left\{i\middle| X_i < Y_{(1)}\right\} + 1$$

$$rang(Y_{(2)}) = \#\left\{i\middle| X_i < Y_{(2)}\right\} + 2$$

$\ldots$

$$rang(Y_{(n)}) = \#\left\{i\middle| X_i < Y_{(n)}\right\} + n$$

und durch Summation finden wir

(3) $$S_Y = N_+ + \frac{n(n+1)}{2},$$

was zu beweisen war.

*Exakte Verteilung von $N_+$ unter der Nullhypothese:*

$X_i, Y_j$ unabhängig je mit gleicher stetiger Verteilungsfunktion $F$;

$$\Rightarrow \quad P\{X_1 < X_2 < \cdots < X_m < Y_1 < \cdots < Y_n\} = \frac{1}{N!}, \text{ wobei } N = m+n;$$

das gleiche gilt für jede andere Permutation der Variablen $X_1,\ldots,X_m,Y_1,\ldots,Y_n$;

$$\Rightarrow \quad P\{X < X < \cdots < X < Y < Y < \cdots < Y\} = \frac{1}{\binom{m+n}{m}} = \frac{m!n!}{(m+n)!};$$

das gleiche gilt für jede andere Kombination der $X$- und $Y$-Variablen.

*Beispiel:* $m=3, n=2: X_1, X_2, X_3, Y_1, Y_2$

Es gibt $\binom{5}{2}=10$ unterscheidbare Kombinationen von *X*- und *Y*-Variablen; jede Kombination hat die Wahrscheinlichkeit $1/10$:

| | Ränge | | | | | $N_+$ |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| 1 | *X* | *X* | *X* | *Y* | *Y* | 6 |
| 2 | *X* | *X* | *Y* | *X* | *Y* | 5 |
| 3 | *X* | *X* | *Y* | *Y* | *X* | 4 |
| 4 | *X* | *Y* | *X* | *X* | *Y* | 4 |
| 5 | *X* | *Y* | *X* | *Y* | *X* | 3 |
| 6 | *X* | *Y* | *Y* | *X* | *X* | 2 |
| 7 | *Y* | *X* | *X* | *X* | *Y* | 3 |
| 8 | *Y* | *X* | *X* | *Y* | *X* | 2 |
| 9 | *Y* | *X* | *Y* | *X* | *X* | 1 |
| 10 | *Y* | *Y* | *X* | *X* | *X* | 0 |

Exakte Verteilung von $N_+$

| $k$ | $P\{N_+ = k\}$ |
|---|---|
| 0 | 1/10 |
| 1 | 1/10 |
| 2 | 2/10 |
| 3 | 2/10 |
| 4 | 2/10 |
| 5 | 1/10 |
| 6 | 1/10 |
| | 1 |

*Testvorschrift* zum Test von Wilcoxon-Mann-Whitney:

Gegeben: $x_1,\ldots,x_m, y_1,\ldots,y_n$

Bestimme rechte kritische Schranke $k_r$ zu $N_+$ unter $H_0$:

$k_r =$ kleinste ganze Zahl so, daß $P_0\{N_+ > k_r\} \le \alpha$.

Bestimme Realisation von $N_+$ bei vorliegender Stichprobe:

$n_+ = \#\{(i,j) \mid x_i < y_j\}$.

Verwirf $H_0$, falls $n_+ > k_r$.

*Testvorschrift mit p-Wert:*

Berechne $p_{crit} = P_0\{N_+ \ge n_+\} = p$-Wert.

Verwirf $H_0$, falls $p_{crit} \le \alpha$.

Es gilt: $p_{crit} \le \alpha \Leftrightarrow n_+ > k_r$, d.h. die beiden Testverfahren sind äquivalent. Die exakte Berechnung des p-Werts unter $H_0$ kann z.B. mit StatXact durchgeführt werden *(Non-Parametrics, Two Independent Samples, Wilcoxon-Mann-Whitney)*.

## 3. Test von Wilcoxon-Mann-Whitney im allgemeinen Fall

Jetzt wollen wir den Test von Wilcoxon-Mann-Whitney im allgemeinen Fall behandeln.

$\left.\begin{matrix} X_1,\ldots,X_m \\ Y_1,\ldots,Y_n \end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases} \text{je mit Verteilungsfunktion } F \\ \text{je mit Verteilungsfunktion } G. \end{cases}$

$H_0: F \equiv G$ (beide Stichproben haben gleiche Verteilung);

$H_1: P\{X < Y\} > P\{X > Y\}$ (*Y*-Werte im Mittel größer als *X*-Werte).

Da die Verteilungsfuktionen *F* und *G* jetzt Unstetigkeitsstellen besitzen können, so müssen wir das Auftreten von Bindungen (ties) berücksichtigen; jede der drei folgenden Bedingungen kann erfüllt sein:

$$P\{X_i = X_j\} > 0 \text{ für } i \neq j;$$
$$P\{X_i = Y_j\} > 0;$$
$$P\{Y_i = Y_j\} > 0 \text{ für } i \neq j.$$

Es sei

$$N_+ = \#\{(i,j) \mid X_i < Y_j\};$$
$$N_0 = \#\{(i,j) \mid X_i = Y_j\};$$
$$N_- = \#\{(i,j) \mid X_i > Y_j\}.$$

Offensichtlich gilt $N_+ + N_- + N_0 = mn =$ Gesamtzahl aller Paare $(i,j)$ mit $i = 1,\ldots,m$ und $j = 1,\ldots,n$. Eine naheliegende Testgröße ist nun

$$U = N_+ + \tfrac{1}{2}N_0 \quad \text{bzw.} \quad \bar{U} = N_- + \tfrac{1}{2}N_0 \qquad (\text{es gilt dann: } U + \bar{U} = mn).$$

Bei der Rangsummenstatistik $S_Y$ von Wilcoxon stellt sich die Frage, wie die Ränge zu berechnen sind beim Auftreten von Bindungen. Das übliche Verfahren besteht darin, dass die Rangzahlen, welche auf eine Bindung fallen, gemittelt werden, und dass dann jede Beobachtung der Bindung diesen Mittelwert als Rangzahl bekommt.

*Beispiel:* $m = 3, n = 2$,

$x_1 = x_2 = y_1 = 7 \; < \; x_3 = 9 \; < \; y_2 = 13$;

die Rangzahlen 1,2,3 fallen auf die Dreierbindung $x_1 = x_2 = y_1$, und somit bekommt jede der drei Beobachtungen $x_1, x_2, y_1$ die Rangzahl $\frac{1}{3}(1+2+3) = 2$.

Wenn nun die Rangsummenstatistik von Wilcoxon definiert wird als

$$S_Y = \sum_{j=1}^{n} rang(Y_j) \qquad (\text{mit Mittelbildung bei Bindungen}),$$

dann gilt analog zu (3) der Zusammenhang

$$S_Y = U + \frac{n(n+1)}{2}$$

(Beweis im Anhang A9). Wir werden im folgenden mit der *U*-Statistik von Mann arbeiten, da sie einfacher und natürlicher ist für unser allgemeines Testproblem. Die naheliegende Testvorschrift wäre nun: Verwirf $H_0$, falls *U* zu groß ist, wobei der kritische Wert für die Statistik *U* aufgrund der exakten Verteilung unter $H_0$ zu bestimmen ist. Im stetigen Fall ist die exakte Verteilung von *U* bzw. $N_+$ unter $H_0$ unabhängig von der speziellen Form der Verteilungsfunktion *F*. Dies ist jedoch im allgemeinen Fall nicht mehr so.

*Beispiel:* Wir betrachen die folgenden beiden Verteilungsfunktionen:

$$F_1(x) = \begin{cases} 0 & \text{für } x < 0 \\ x & \text{für } 0 \le x \le 1 \\ 1 & \text{für } x > 1 \end{cases} \qquad F_2(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{2} & \text{für } x = 0 \\ \frac{1}{2}(1+x) & \text{für } 0 < x \le 1 \\ 1 & \text{für } x > 1. \end{cases}$$

$F_1$ ist die Verteilungsfunktion zur Gleichverteilung im Intervall $[0,1]$, während die Verteilungsfunktion $F_2$ einen Sprung von der Höhe $1/2$ im Punkt $x = 0$ besitzt. Bei $F_1$ werden keine Bindungen auftreten, während bei $F_2$ etwa die Hälfte aller Beobachtungen auf den Punkt $x = 0$ fällt. Die Verteilung der Statistik *U* wird also bei den beiden Verteilungen verschieden sein!

Die Lösung unseres Problems besteht darin, dass man die *bedingte Verteilung* von *U bei gegebener Bindungskonfiguration* berechnet.

*Beispiel:* $m=3, n=2$, $x_1=x_2=y_1=7 < x_3=9 < y_2=13$;
die *Bindungskonfiguration* ist gegeben durch $(3,1,1)$, weil zunächst eine Dreier-Bindung, und dann zwei einzelne (ungebundene) Beobachtungen vorliegen.

Eine *Bindungskonfiguration* ist eine Zerlegung von $N=m+n$ in eine Summe von natürlichen Zahlen. Für $N=5$ ergeben sich die folgenden Bindungskonfigurationen:

1) (1,1,1,1,1) (keine Bindungen)
2) (2,1,1,1)
3) (1,2,1,1)
…
16) (5) (alle fünf Beobachtungen haben den gleichen Wert).

Die Anzahl der möglichen Bindungskonfigurationen bei gegebenem *N* beträgt $2^{N-1}$ (Gleichheitszeichen bzw. Ungleichheitszeichen zwischen die *N* Beobachtungen setzen).

*Exakte Verteilung von U unter der Nullhypothese bei gegebener Bindungskonfiguration:*

$X_i, Y_j$ unabhängig je mit gleicher stetiger Verteilungsfunktion $F$;
$b=(b_1,b_2,\ldots)$ gegebene Bindungskonfiguration, $b_1+b_2+\cdots=N$;
jede Anordnung der *X*- und *Y*-Variablen auf die *N* Plätze besitzt die Wahrscheinlichkeit $1\Big/\binom{N}{m}$;
dies ermöglicht die Bestimmung der exakten bedingten Verteilung.

*Beispiel:* $m=3, n=2, N=m+n=5; b=(3,1,1)$
Es gibt $\binom{5}{2}=10$ unterscheidbare Anordnungen der *X*- und *Y*-Variablen auf die 5 Plätze;
jede Anordnung hat die Wahrscheinlichkeit $1/10$:

| | Rangzahlen | | | | | *U* |
|---|---|---|---|---|---|---|
| | 2 | 2 | 2 | 4 | 5 | |
| 1 | *X* | *X* | *X* | *Y* | *Y* | 6 |
| 2 | *X* | *X* | *Y* | *X* | *Y* | 4 |
| 3 | *X* | *X* | *Y* | *Y* | *X* | 3 |
| 4 | *X* | *Y* | *X* | *X* | *Y* | 4 |
| 5 | *X* | *Y* | *X* | *Y* | *X* | 3 |
| 6 | *X* | *Y* | *Y* | *X* | *X* | 1 |
| 7 | *Y* | *X* | *X* | *X* | *Y* | 4 |
| 8 | *Y* | *X* | *X* | *Y* | *X* | 3 |
| 9 | *Y* | *X* | *Y* | *X* | *X* | 1 |
| 10 | *Y* | *Y* | *X* | *X* | *X* | 1 |

Exakte Verteilung von *U* gegeben $b=(3,1,1)$

| *k* | $P\{U=k\}$ |
|---|---|
| 1 | 3/10 |
| 3 | 3/10 |
| 4 | 3/10 |
| 6 | 1/10 |
| | 1 |

Wir sind also in der Lage, *bei gegebener Bindungskonfiguration* die exakte Verteilung von *U* unter $H_0$ zu berechnen; diese Verteilung hängt nicht ab von der unbekannten wahren Verteilungsfunktion *F*. Somit können wir für jede Bindungskonfiguration $b=(b_1,b_2,\ldots)$ mit $b_1+b_2+\cdots=N$ einen korrekten bedingten Test durchführen, für welchen gilt.

$$P_0\{\text{Test verwirft } H_0 \mid b\} \le \alpha .$$

Es sei $B_N = \{b = (b_1, b_2, \ldots) | b_1 + b_2 + \cdots = N\}$ = Menge aller Bindungskonfig. zum Umfang *N*. Dann gilt:

$$P_0\{\text{Test verwirft } H_0\} = \sum_{b \in B_N} \underbrace{P_0\{(X,Y) \to b\}}_{\text{abhängig von } F} \underbrace{P_0\{\text{Test verwirft } H_0 | b\}}_{\leq \alpha \text{ für jedes } b \in B_N}$$

$$\leq \alpha \underbrace{\sum_{b \in B_N} P_0\{(X,Y) \to b\}}_{=1} = \alpha,$$

d.h. bei diesem Testverfahren ist nicht nur die bedingte Irrtumswahrscheinlichkeit 1. Art, sondern auch die globale (unbedingte) Irrtumswahrscheihnlichkeit 1. Art $\leq \alpha$. Es handelt sich also um einen korrekten, exakten Test zum Signifikanzniveau $\alpha$, der keinerlei Voraussetzungen über die unbekannte Verteilungsfunktion *F* benötigt.

*Zusammenfassung:*

$\left.\begin{matrix} X_1, \ldots, X_m \\ Y_1, \ldots, Y_n \end{matrix}\right\}$ unabhängige Zufallsgrößen $\begin{cases} \text{je mit Verteilungsfunktion } F \\ \text{je mit Verteilungsfunktion } G; \end{cases}$

$x_1, \ldots, x_m, y_1, \ldots, y_n$ zugehörige Realisationen;

$H_0: F \equiv G$ (beide Stichproben haben gleiche Verteilung);

$H_1: P\{X < Y\} > P\{X > Y\}$ (*Y*-Werte im Mittel größer als *X*-Werte);

$N_+ = \#\{(i,j) | X_i < Y_j\}$

$N_0 = \#\{(i,j) | X_i = Y_j\}$

$U = N_+ + \frac{1}{2} N_0$

$x_1, \ldots, x_m, y_1, \ldots, y_n \rightarrow b = (b_1, b_2, \ldots) =$ Bindungskonfiguration zur vorliegenden Stichprobe;

exakte Verteilung von *U* gegeben $b = (b_1, b_2, \ldots)$ unter $H_0$ berechenbar;

berechne $p_{crit} = P_0\{U \geq u_{obs} | b\}$, wobei $u_{obs}$ der Wert von *U* bei der vorliegenden Stichprobe ist;

verwirf $H_0$, falls $p_{crit} \leq \alpha$.

Dann gilt: $P_0\{\text{Test verwirft } H_0\} \leq \alpha$ d.h. das (unbedingte) Signifikanzniveau $\alpha$ wird eingehalten.

*Bemerkungen:*

1) Keinerlei Voraussetzungen an die Verteilungsfunktionen *F* und *G*, auch keine Stetigkeit.
2) Die exakte bedingte Verteilung ist datenabhängig (abhängig von Bindungskonfiguration der vorliegenden Stichprobe); es wird jedoch nicht nur das bedingte, sondern auch das globale (unbedingte) Signifikanzniveau eingehalten.
3) Zur praktischen Durchführung ist ein Computerprogramm notwendig (StatXact, SAS, SPSS).
4) Falls die Berechnung der exakten bedingten Verteilung zu schwierig wird, so sind Mone-Carlo-Verfahren verfügbar.
5) Der Test ist konsistent, d.h. wenn die Alternative vorliegt konvergiert die Ablehnwahrscheinlichkeit für wachsende Stichprobenumfänge gegen 1.

# 9. Lineare Rangtests beim Einstichprobenproblem

## 1. Problemstellung

A) *Klassischer t-Test*

Es seien $X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit einer Normalverteilung $N(\mu,\sigma^2)$ mit unbekanntem $\mu$ und $\sigma$. Wir betrachten das Testproblem

$H_0: \mu \le \mu_0$ gegen $H_1: \mu > \mu_0$.

Beim klassischen t-Test betrachtet man die Testgröße

$$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n}, \text{ wobei } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 ;$$

Falls $\mu = \mu_0$, so gilt $T \sim t(n-1)$, d.h. die Zufallsgröße $T$ besitzt dann eine t-Verteilung mit $n-1$ Freiheitsgraden. Die Testvorschrift des t-Tests zum Signifikanzniveau $\alpha$ lautet daher: Verwirf $H_0$, falls $T > t_{1-\alpha}(n-1)$.

B) *Symmetrisches Lagemodell*

Wir wollen die Problemstellung des klassischen t-Tests verallgemeinern und dabei auf die Normalverteilungsannahme verzichten. Wir betrachten das *symmetrische Lagemodell*:

$F$ sei eine Verteilungsfunktion mit einer symmetrischen Dichte $f$, d.h. $f(-x) = f(x)$ für $x \in \mathbb{R}$;
$F_\mu(x) = F(x-\mu)$, $f_\mu(x) = f(x-\mu)$, $\mu \in \mathbb{R}$, $\mu$ = med($F$) = Lageparameter (Verschiebung um $\mu$).

Nun seien $X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion $F_\mu$. Die Verteilungsfunktion $F$ und der Parameter $\mu$ seien unbekannt. Wir betrachten das Testproblem

$H_0: \mu \le \mu_0$ gegen $H_1: \mu > \mu_0$.

Falls $F$ die Verteilungsfunktion zu $N(0,\sigma^2)$ ist, so haben wir wiederum die Problemstellung des klassischen t-Tests. In diesem Kapitel interessieren wir uns für Testverfahren, welche das allgemeinere Problem mit dem symmetrischen Lagemodell lösen, und zwar für *lineare Rangtests*. Wir wollen die Grundidee dieser Testverfahren beschreiben am Beispiel des Vorzeichentests und des Einstichprobentests von Wilcoxon.

*Vorzeichentest*

Es sei $D_i = X_i - \mu_0$ für $i = 1,\ldots,n$ und

$N_+ = \#\{i \mid X_i > \mu_0\} = \#\{i \mid D_i > 0\}$ = Anzahl der positiven Differenzen $D_i$,

$N_- = \#\{i \mid X_i < \mu_0\} = \#\{i \mid D_i < 0\}$ = Anzahl der negativen Differenzen $D_i$.

Da die Zufallsgrößen $X_i$ hier eine stetige Verteilungsfunktion besitzen, brauchen wir keine Bindungen zu berücksichtigen. Der Vorzeichentest verwirft $H_0$, falls $N_+$ zu groß, d.h. falls $N_-$ zu klein ist. Falls $\mu = \mu_0$, so ist $P\{X_i > \mu_0\} = P\{D_i > 0\} = \frac{1}{2}$, und damit besitzt dann $N_+$ eine Binomialverteilung $Bi(n,\frac{1}{2})$ unabhängig von der speziellen Form der Verteilungsfunktion $F$.

*Einstichprobentest von Wilcoxon*

Wiederum sei $D_i = X_i - \mu_0$ für $i = 1,\ldots,n$. Beim Vorzeichentest wurde nur die Anzahl der positiven und negativen Differenzen $D_i$ gezählt. Jetzt sollen diese Differenzen gewichtet werden:

$|D_i|$ groß $\Rightarrow$ großes Gewicht;

$|D_i|$ klein $\Rightarrow$ kleines Gewicht.

Eine einfache Gewichtung besteht darin, dass man die Differenzen $|D_i|$ ordnet nach ihrer Größe und als Gewichte die Rangzahlen $R_i$ wählt:

$R_i = rang(|D_i|) =$ Platznummer von $D_i$ in der geordneten Liste.

Anstelle der Anzahl $N_+$ der positiven Differenzen wird jetze die Rangsumme der positiven Differenzen als Testgröße gewählt. Wir setzen

$$V_i = \begin{cases} 1 & \text{falls } D_i > 0 \\ 0 & \text{falls } D_i < 0 \end{cases}$$

und definieren unsere Testgrößen als

$$L_+ = \sum_{i=1}^{n} V_i R_i = \text{ Rangsumme der positiven Differenzen } D_i;$$

$$L_- = \sum_{i=1}^{n} (1 - V_i) R_i = \text{ Rangsumme der negativen Differenzen } D_i.$$

Es gilt $L_+ + L_- = n$, da ja keine Bindungen auftreten können. Beim Vorzeichentest haben alle Differenzen das gleiche Gewicht 1, und die beiden Testgrößen $N_+$ und $N_-$ können auch geschrieben werden als

$$N_+ = \sum_{i=1}^{n} V_i = \text{ Anzahl der positiven Differenzen } D_i;$$

$$N_- = \sum_{i=1}^{n} (1 - V_i) = \text{ Anzahl der negativen Differenzen } D_i.$$

Die natürliche Testidee mit der Statistik $L_+$ bzw. $L_-$ ist die folgende:

$L_+$ groß $\rightarrow$ $H_1$,

$L_+$ klein $\rightarrow$ $H_0$.

Es bleibt die entscheidende Frage, wie der kritische Wert zu bestimmen ist ohne Kenntnis der Verteilungsfunktion *F*. Falls $\mu = \mu_0$, so besitzen die Zufallsgrößen $D_i = X_i - \mu_0$ alle die symmetrische Dichte *f*; positive und negative Werte von $D_i$ sind gleich verteilt. Nun gilt aus Symmetriegründen

$$P\{D_1 > 0, R_1 = j\} = P\{D_1 < 0, R_1 = j\} \text{ für } j = 1, \ldots, n,$$

und daraus folgt

(1) $$P\{D_1 > 0, R_1 = j\} = \tfrac{1}{2} P\{R_1 = j\} \text{ für } j = 1, \ldots, n.$$

Nun ist $\{D_1 > 0\} = \{V_1 = 1\}$, und daher folgt aus (1)

$$P\{V_1 = 1, R_1 = j\} = \tfrac{1}{2} P\{R_1 = j\} = P\{V_1 = 1\} P\{R_1 = j\} \text{ für } j = 1, \ldots, n,$$

und dies bedeutet, dass die Zufallsgrößen $V_1$ und $R_1$ stochastisch unabhängig sind; ebenso zeigt man, dass $V_i$ und $R_i$ unabhängig sind für $i = 1, \ldots, n$. Falls $\mu = \mu_0$, so werden also die „Vorzeichen“ $V_1, \ldots, V_n$ zufällig auf die Rangzahlen $R_1, \ldots, R_n$ verteilt, und damit läßt sich die exakte Verteilung von $L_+$ bzw. $L_-$ mit kombinatorischen Methoden bestimmen unabhängig von der speziellen Form der unbekannten Verteilungsfunktion *F* (vgl. nachfolgendes Beispiel).

*Beispiel:* $n=4$, $D_i = X_i - \mu_0$, $R_i = rang\left(|D_i|\right)$;

| | $rang\left(\|D_i\|\right)$ | | | | $L_+$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| 1 | + | + | + | + | 10 |
| 2 | – | + | + | + | 9 |
| 3 | + | – | + | + | 8 |
| 4 | + | + | – | + | 7 |
| 5 | + | + | + | – | 6 |
| 6 | – | – | + | + | 7 |
| 7 | – | + | – | + | 6 |
| 8 | – | + | + | – | 5 |
| 9 | + | – | – | + | 5 |
| 10 | + | – | + | – | 4 |
| 11 | + | + | – | – | 3 |
| 12 | – | – | – | + | 4 |
| 13 | – | – | + | – | 3 |
| 14 | – | + | – | – | 2 |
| 15 | + | – | – | – | 1 |
| 16 | – | – | – | – | 0 |

Exakte Verteilung von $L_+$ für $\mu = \mu_0$

| $k$ | $P\{L_+ = k\}$ |
|---|---|
| 0 | 1/16 |
| 1 | 1/16 |
| 2 | 1/16 |
| 3 | 2/16 |
| 4 | 2/16 |
| 5 | 2/16 |
| 6 | 2/16 |
| 7 | 2/16 |
| 8 | 1/16 |
| 9 | 1/16 |
| 10 | 1/16 |
| $\sum$ | 1 |

Für $\mu = \mu_0$ besitzt jede der $2^4 = 16$ Vorzeichenkombinationen die Wahrscheinlichkeit $1/16$.

## 2. Allgemeine lineare Rangtests

$X_1, \ldots, X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion $F_\mu$ mit $F_\mu(x) = F(x-\mu)$;
$f$ Dichte zu $F$; $f$ symmetrisch bzgl. 0, d.h. $f(-x) = f(x)$ für $x \in \mathbb{R}$;
$H_0: \mu \le 0$ gegen $H_1: \mu > 0$ (wir nehmen jetzt an, dass $\mu_0 = 0$, sodass $D_i = X_i - \mu_0 = X_i$);
$R_i = rang(|X_i|)$ (Beobachtungen ordnen nach Größe von $|X_i|$);

$$V_i = \begin{cases} 1 & \text{falls } X_i > 0 \\ 0 & \text{falls } X_i < 0 \end{cases}$$

$w_1, \ldots . w_n$ Gewichte;
$gew(i) = w_i =$ Gewicht (Score) bei Rangzahl $i$:

$$L_+ = \sum_{i=1}^{n} V_i \, gew(R_i) = \sum_{i=1}^{n} V_i \, w_{R_i} = \text{ Gewichtssumme der positiven Beobachtungen } X_i .$$

Vorzeichentest: $w_i = 1$ für $i = 1, \ldots, n$;
Test von Wilcoxon: $w_i = i$ für $i = 1, \ldots, n$.

Falls $\mu = 0$, so sind die Vorzeichenvariablen $V_1, \ldots, V_n$ unabhängige symmetrische Zweipunktvariablen mit $P\{V_i = 1\} = P\{V_i = 0\} = 1/2$ für $i = 1, \ldots, n$. Weiter sind die Vektoren $(V_1, \ldots, V_n)$ und $(R_1, \ldots, R_n)$ stochastisch unabhängig (vgl. Abschnitt 1). Die exakte Verteilung von $L_+$ kann also auch bei allgemeinen linearen Rangtest wie beim Test von Wilcoxon (vgl. Abschnitt 1) mit kombinatorischen Methoden bestimmt werden; die Verteilung ist unabhängig von der speziellen Form der unbekannten Verteilungsfunktion $F$.

*Erwartungswert und Varianz von* $L_+$ *falls* $\mu = 0$*:*

Wir wollen in der Statistik $L_+$ die Summationsreihenfolge ändern, damit unsere Berechnungen einfacher werden. Wir können schreiben

$$L_+ = \sum_{i=1}^{n} V_i\, g(R_i) = \sum_{j=1}^{n} W_j\, g(j),$$

wobei

$$W_j = \begin{cases} 1 & \text{falls die Beobachtung } X_i \text{ mit Rangzahl } j \text{ ein positives Vorzeichen hat} \\ 0 & \text{falls die Beobachtung } X_i \text{ mit Rangzahl } j \text{ ein negatives Vorzeichen hat.} \end{cases}$$

Die Zufallsgrößen $W_1,\ldots,W_n$ sind stochastisch unabhängig und gleich verteilt wie $V_1,\ldots,V_n$. Daher gilt

$$E_0(L_+) = E_0\left(\sum_{j=1}^{n} W_j\, g(j)\right) = \tfrac{1}{2}\sum_{j=1}^{n} g(j)$$

$$Var_0(L_+) = Var_0\left(\sum_{j=1}^{n} W_j\, g(j)\right) = \sum_{j=1}^{n} g^2(j)\, Var_0(W_j) = \tfrac{1}{4}\sum_{j=1}^{n} g^2(j)$$

## 3. Andere Darstellung von $L_+$ beim Test von Wilcoxon

$X_1,\ldots,X_n$ ursprüngliche Zufallsvariablen;

$R_1,\ldots,R_n$ Rangzahlen, $R_i = rang(|X_i|)$;

$V_1,\ldots,V_n$ Vorzeichenvariablen, $V_i = \begin{cases} 1 & \text{falls } X_i > 0 \\ 0 & \text{falls } X_i < 0; \end{cases}$

$\left.\begin{matrix} x_1,\ldots,x_n \\ r_1,\ldots,r_n \\ v_1,\ldots,v_n \end{matrix}\right\}$ Realisationen zu den entsprechenden Zufallsgrößen;

$L_+ = \sum_{i=1}^{n} V_i R_i$ Statistik zum Einstichprobentest von Wilcoxon;

$l_+ = \sum_{i=1}^{n} v_i r_i$ zugehörige Realisation.

Wir können annehmen, dass $x_1 < x_2 < \cdots < x_k$ (umnumerieren!), und es sei dann $x_k < 0$ und $x_{k+1} > 0$. Dann ist

$$l_+ = \sum_{i=1}^{n} v_i r_i = \sum_{i=k+1}^{n} r_i\,.$$

Nun gilt

$$r_{k+1} = \#\left\{i \,\middle|\, |x_i| \le x_{k+1}\right\} = \#\left\{i \,\middle|\, i \le k+1,\, x_i + x_{k+1} > 0\right\}$$

$$r_{k+2} = \#\left\{i \,\middle|\, |x_i| \le x_{k+2}\right\} = \#\left\{i \,\middle|\, i \le k+2,\, x_i + x_{k+2} > 0\right\}$$

…

und daher ist

$$l_+ = \sum_{i=k+1}^{n} r_i = \#\left\{(i,j) \,\middle|\, i \le j,\, x_i + x_j > 0\right\}.$$

Da dies für beliebige Stichprobenwerte gilt, so haben wir die Darstellung

$$L_+ = \sum_{i=1}^{n} V_i R_i = \#\left\{(i,j) \middle| i \le j,\ X_i + X_j > 0\right\}.$$

Diese Darstellung zeigt, dass der Test von Wilcoxon verwandt ist mit dem sogenannten Hodges-Lehmann-Schätzer für $\mu$:

$$\hat{\mu} = Ze\left\{\frac{X_i + X_j}{2}, i \le j\right\}$$

(Ze = Zentralwert). Der Punktschätzer $\hat{\mu}$ ist die duale Schätzfunktion zum Einstichprobentest von Wilcoxon, wie etwa das arithmetische Mittel $\bar{X}$ die duale Schätzfunktion zum einfachen t-Test ist.

Wie bei den linearen Rangtests zum Zweistichprobenproblem (vgl. Kapitel 7), interessieren wir uns für die folgenden Fragen.

a) Vergleich des t-Tests mit dem Test von Wilcoxon?
b) Welches ist die optimale Gewichtung bei gegebener Dichte *f*?
c) Asymptotische Gütefunktionen, asymptotische relative Effizienz?

## 4. Asymptotische Gütefunktion beim einfachen t-Test

$X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit einer Normalverteilung $N(\vartheta,\sigma^2)$;
$H_0: \vartheta \le 0$ gegen $H_1: \vartheta > 0$ ($\vartheta$ ist der Lageparameter im Lagemodell);
$T = \frac{\bar{X}}{S}\sqrt{n}$ mit $S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$;
t-Test: Verwirf $H_0$, falls $T > t_{1-\alpha}(n-1)$.

Wir wollen die asymptotische Gütefunktion des t-Tests bestimmen. Dazu führen wir den Parameter $\delta = \vartheta\sqrt{n}$ ein und schreiben die Gütefunktion als

$$\beta_n(\delta) = P\{\text{Test verwirft } H_0\} = P\{T > t_{1-\alpha}(n-1)\}.$$

Für $n \to \infty$ ($\delta, \alpha$ fest) gilt

(2) $\beta_n(\delta) \longrightarrow P\left\{N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sigma}\right\}$

(vgl. Abschnitt 2.2).

## 5. Asymptotische Gütefunktion bei allgemeinen linearen Rangtests

$X_1,\ldots,X_n$ unabhängige Zufallsgrößen je mit der Dichte $f_\vartheta(x) = f(x-\vartheta)$;
$f$ symmetrisch, d.h. $f(-x) = f(x)$ für $x \in \mathbb{R}$;
$H_0: \vartheta \le 0$ gegen $H_1: \vartheta > 0$ ($\vartheta$ ist der Lageparameter im Lagemodell);
$R_1,\ldots,R_n$ Rangzahlen, $R_i = rang(|X_i|)$;
$V_1,\ldots,V_n$ Vorzeichenvariablen, $V_i = \begin{cases} 1 & \text{falls } X_i > 0 \\ 0 & \text{falls } X_i < 0; \end{cases}$
$w: [0,1] \to \mathbb{R}$ gewichtserzeugende Funktion, $w(u) \ge 0$;
$gew(i) = w_i = w\left(\frac{i}{n+1}\right) =$ Gewicht zur Rangzahl $i$, $i = 1,\ldots,n$;

$$L_+ = \sum_{i=1}^{n} V_i\, gew(R_i) = \sum_{i=1}^{n} V_i\, w_{R_i} = \sum_{i=1}^{n} V_i\, w\left(\frac{R_i}{n+1}\right)$$ Teststatistik zu linearem Rangtest;

$$\bar{L}_+ = \frac{1}{n} L_+ = \frac{1}{n} \sum_{i=1}^{n} V_i \, w\left(\frac{R_i}{n+1}\right).$$

Falls $\vartheta = 0$, so gilt

$$E_0\left(L_+\right) = \tfrac{1}{2} \sum_{j=1}^{n} w_j = \tfrac{1}{2} \sum_{j=1}^{n} w\left(\frac{j}{n+1}\right)$$

$$Var_0\left(L_+\right) = \tfrac{1}{4} \sum_{j=1}^{n} w_j^2 = \tfrac{1}{4} \sum_{j=1}^{n} w^2\left(\frac{j}{n+1}\right)$$

und somit erhalten wir für $n \to \infty$

$$\text{(3)} \qquad \begin{aligned} \mu_0 = E_0\left(\bar{L}_+\right) &= \frac{1}{2n} \sum_{j=1}^{n} w\left(\frac{j}{n+1}\right) \quad \longrightarrow \quad \tfrac{1}{2} \int_0^1 w(u)\,du; \\ n\sigma_0^2 = n Var_0\left(\bar{L}_+\right) &= \frac{1}{4n} \sum_{j=1}^{n} w^2\left(\frac{j}{n+1}\right) \quad \longrightarrow \quad \tfrac{1}{4} \int_0^1 w^2(u)\,du. \end{aligned}$$

Der lineare Rangtest mit der Teststatistik $\bar{L}_+$ verwirft $H_0$, falls $\bar{L}_+ > k_r$, wobei $k_r$ die rechte kritische $\alpha$-Schranke der exakten Verteilung von $\bar{L}_+$ unter der Annahme $\vartheta = 0$ bezeichnet. Nun wollen wir die asymptotische Gütefunktion dieses Tests bestimmen. Es sei $\delta = \vartheta\sqrt{n}$ und wir betrachten den Grenzübergang $n \to \infty$ mit $\delta, \alpha$ fest. Die Gütefunktion

$$\text{(4)} \quad \beta_n(\delta) = P\left\{\text{Test verwirft } H_0\right\} = P\left\{\bar{L}_+ > k_r\right\} = P\left\{\frac{\bar{L}_+ - \mu}{\sigma} > \frac{k_r - \mu}{\sigma}\right\},$$

wobei $\mu = E(\bar{L}_+)$ und $\sigma^2 = Var(\bar{L}_+)$. (Wir haben im vorangehenden Abschnitt mit $\sigma^2$ die Varianz der normalverteilten Zufallsgrößen $X_i$ bezeichnet; diese Varianz kommt hier nicht vor.) Falls

$$\int w(u)du < \infty \quad \text{und} \quad \int w^2(u)du < \infty,$$

so können die Einzelgewichte $w_1, \ldots, w_n$ nicht allzu extrem variieren, und es ist plausibel, dass dann bei unserem Grenzübergang der folgende Grenzwertsatz gilt:

$$\text{(5)} \quad \frac{\bar{L}_+ - \mu}{\sigma} \xrightarrow{\;d\;} N(0,1)$$

(vgl. Hettmansperger, 1984, section 2.8). Nun ist

$$\text{(6)} \quad \frac{k_r - \mu}{\sigma} = \left(\frac{k_r - \mu_0}{\sigma_0} - \frac{\mu - \mu_0}{\sigma_0}\right)\frac{\sigma_0}{\sigma},$$

und es ist plausibel, dass

$$\frac{k_r - \mu_0}{\sigma_0} \to z_{1-\alpha} \qquad \text{aufgrund des Grenzwertsatzes (5),}$$

$$\frac{\sigma_0}{\sigma} \to 1 \qquad \text{denn } \vartheta = \delta/\sqrt{n} \to 0.$$

Somit haben wir in (6) noch den Grenzwert von $(\mu - \mu_0)/\sigma_0$ zu bestimmen. Wir schreiben

$$\frac{\mu - \mu_0}{\sigma_0} = \frac{\vartheta}{\sigma_0} \times \frac{\mu - \mu_0}{\vartheta} = T_1 \times T_2.$$

Bei unserem Grenzübergang gilt aufgrund von (3)

$$T_1 = \frac{\vartheta}{\sigma_0} = \frac{\delta}{\sigma_0\sqrt{n}} \longrightarrow \frac{\delta}{\tau}, \quad \text{wobei } \tau^2 = \tfrac{1}{4} \int_0^1 w^2(u)\,du,$$

und im Anhang A10 wird gezeigt, dass

$$T_2 = \frac{\mu - \mu_0}{\vartheta} \longrightarrow I_f \ ,$$

wobei

$$\begin{aligned} I_f \ &= \ 2\int_0^\infty w'\big[2F(x)-1\big]\, f^2(x)\, dx \ + \ w(0)\, f(0) \ = \ -\int_0^\infty w\big[2F(x)-1\big]\, f'(x)\, dx \\ &= \int_0^1 w'(u)\, f\left[F^{-1}\left(\frac{u+1}{2}\right)\right] du \ + w(0)\, f(0) \ = \ \tfrac{1}{2}\int_0^1 w(u)\, w_f(u)\, du \end{aligned} \tag{7}$$

mit

$$w_f(u) = -\frac{f'\left[F^{-1}\left(\frac{u+1}{2}\right)\right]}{f\left[F^{-1}\left(\frac{u+1}{2}\right)\right]}.$$

Dabei benötigen wir die folgenden Voraussetzungen

$f(x) > 0$ für alle $x \in \mathbb{R}$ ,

Dichte $f$ und Gewichtsfunktion $w$ differenzierbar,

$w\big[2F(x)-1\big]\, f(x) \to 0$ für $x \to \infty$ .

Somit gilt beim Grenzübergang $n \to \infty$ ($\delta, \alpha$ fest)

$$\frac{\mu - \mu_0}{\sigma_0} = T_1 \times T_2 \longrightarrow \frac{\delta}{\tau} \times I_f \ .$$

Wir erhalten also als Grenzwert der Gütefunktion (4) aufgrund von (5) und (6)

$$\beta_n(\delta) = P\left\{\frac{\bar{L}_+ - \mu}{\sigma} > \frac{k_r - \mu}{\sigma}\right\} \ \to \ \beta(\delta) \ = \ P\big\{N(0,1) > z_{1-\alpha} - c\,\delta\big\} \ = \ P\big\{N(0,1) < -z_{1-\alpha} + c\,\delta\big\},$$

wobei

$$c = \frac{I_f}{\tau} = \frac{\int_0^1 w(u)\, w_f(u)\, du}{\left(\int_0^1 w^2(u)\, du\right)^{\frac{1}{2}}} \quad \text{mit} \quad w_f(u) = -\frac{f'\left[F^{-1}\left(\frac{u+1}{2}\right)\right]}{f\left[F^{-1}\left(\frac{u+1}{2}\right)\right]}. \tag{8}$$

Beachte, dass hier im Unterschied zum Zweistichprobenproblem gilt

$$\int_0^1 w_f(u)\, du = 2\int_0^\infty f'(x)\, dx = 2\, f(0)$$

(Substitution: $(u+1)/2 = F(x)$ ). Damit bleibt die Konstante $c$ invariant bei einer multiplikativen Transformation der Gewichte $w(u) \ \to \ a\, w(u)$ , jedoch nicht bei einer allgemeinen Lineartransformation $w(u) \ \to \ a\, w(u) + b$ wie beim Zweistichprobenproblem. Bei einer Skalenänderung $f(x) \ \to \ f(x/a)/a$ gilt $f'(x) \ \to \ f'(x/a)\big/a^2$ und somit $w_f(u) \ \to \ w_f(u)\big/a$ ; damit ergibt sich $c \ \to \ c/a$ . Beim doppelten t-Test ändert sich die Konstante $c$ auf die gleiche Weise, und damit bleibt die relative Effizienz beim Vergleich mit einem linearen Rangtest invariant bei einer Skalenänderung.

*Frage:*
Wie ist bei gegebenen Dichte $f(x)$ die Gewichtsfunktion $w(u)$ zu wählen, damit der lineare Rangtest möglichst trennscharf wird?

*Antwort:*
Die Gewichtsfunktion ist so zu wählen, dass die Konstante *c* maximal wird. Dies ist der Fall, wenn $w(u) = w_f(u)$. Es ist dann $c = \left(\int_0^1 w_f^2(u)\,du\right)^{\frac{1}{2}}$. Bei gegebener Dichtefunktion *f* besitzt also der lineare Rangtest mit der Gewichtsfunktion $w_f(u)$ unter allen linearen Rangtests die bestmögliche asymptotische Gütefunktion (Beweis im Anhang A11).

## 6. Asymptotische relative Effizienz, Beispiele

In diesem Abschnitt wollen wir die asymptotische relative Effizienz von einigen linearen Rangtests im Vergleich zum einfachen t-Test untersuchen. Wir haben sowohl beim einfachen t-Test als auch bei den linearen Rangtests den gleichen Grenzübergang durchgeführt und dabei jeweils Grenzfunktionen von der folgenden Form gefunden:

$$\beta_n(\delta) \longrightarrow P\{N(0,1) < -z_{1-\alpha} + c\,\delta\},$$

wobei die Konstante *c* abhängig ist vom jeweiligen Test. Beim t-Test ist gemäß (2) $c = c_t = 1/\sigma$, und bei den linearen Rangtests ist die Konstante *c* gegeben durch (8), wobei das Integral $I_f$ gemäß (7) auf unterschiedliche Weise berechnet werden kann.

a) *Test von Wilcoxon*
Hier ist

$$w(u) = u;$$

$$w_i = w\left(\frac{i}{n+1}\right) = \frac{i}{n+1} \quad \text{für } i = 1,\ldots,n \qquad (\text{äquivalent zu } w_i = i);$$

$$\tau^2 = \tfrac{1}{4}\int_0^1 w^2(u)\,du = \tfrac{1}{4}\int_0^1 u^2\,du = \tfrac{1}{12};$$

$$c = c_W = \frac{I_f}{\tau} = 2\sqrt{12}\int_0^\infty w'\big(2F(x)-1\big)\,f^2(x)\,dx + \sqrt{12}\,w(0)\,f(0)$$

$$= 2\sqrt{12}\int_0^\infty f^2(x)\,dx = \sqrt{12}\int_{-\infty}^{\infty} f^2(x)\,dx.$$

Die asymptotische relative Effizienz des Tests von Wilcoxon im Vergleich zum einfachen t-Test beträgt also

$$are(W:t) = \left(\frac{c_W}{c_t}\right)^2 = 12\sigma^2\left(\int f^2(x)\,dx\right)^2,$$

wobei $\sigma^2 = Var(X_i)$. Wir haben hier genau das gleiche Ergebnis wie beim Zweistichprobenproblem. Es gilt $are(W:t) \geq 108/125 = 0.864$ (wie beim Zweistichprobenproblem). Die asymptotische relative Effizienz des Tests von Wilcoxon im Vergleich zum t-Test kann beliebig gross werden ($\sigma^2 \to \infty$), sie kann aber nie schlechter werden als 0.864!

Für die logistische Verteilung ist

$$F(x)=\frac{1}{1+e^{-x}},\quad f(x)=\frac{e^{-x}}{(1+e^{-x})^2},\quad -\frac{f'(x)}{f(x)}=\frac{1-e^{-x}}{1+e^{-x}}=2F(x)-1;$$

es gilt

$$x=F^{-1}\left(\frac{u+1}{2}\right)\ \Leftrightarrow\ F(x)=\frac{u+1}{2}\ \Leftrightarrow\ 2F(x)-1=u\,,$$

und wir erhalten

$$w_f(u)=-\frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}=u\,.$$

Dies ist die Gewichtsfunktion des Tests von Wilcoxon, und somit ist der Test von Wilcoxon der optimale lineare Rangtest bei der logistischen Verteilung (wie beim Zweistichprobenproblem).

b) *Normal-Scores-Test (Test mit Normal-Gewichten)*

Hier ist

$$w(u)=\Phi^{-1}\left(\frac{u+1}{2}\right),\ \text{wobei}\ \Phi=\ \text{Verteilungsfunktion zu}\ N(0,1);$$

$$w_i=w\left(\frac{i}{n+1}\right)=\Phi^{-1}\left(\tfrac{1}{2}+\frac{i}{2(n+1)}\right)\ \text{für}\ i=1,\ldots,n;$$

$$\tau^2=\tfrac{1}{4}\int_0^1 w^2(u)\,du=\tfrac{1}{4}\int_0^1\left(\Phi^{-1}\left(\frac{u+1}{2}\right)\right)^2 du=\tfrac{1}{2}\int_0^\infty z^2\varphi(z)\,dz=\tfrac{1}{4};$$

$$c=c_{NS}=\frac{I_f}{\tau}=2I_f\ =\int_0^1 w(u)\,w_f(u)\,du=\int_0^1\Phi^{-1}\left(\frac{u+1}{2}\right)w_f(u)\,du.$$

Die asymptotische relative Effizienz des Normal-Scores-Tests im Vergleich zum t-Test beträgt

$$are(NS:t)=\left(\frac{c_{NS}}{c_t}\right)^2=\sigma^2c_{NS}^2=\sigma^2\left(\int_0^1\Phi^{-1}\left(\frac{u+1}{2}\right)w_f(u)\,du\right)^2.$$

Im Spezialfall der Normalverteilung ist $f=\varphi$, und da $\varphi'(z)/\varphi(z)=-z$, so gilt dann $w_f(u)=\Phi^{-1}\big((u+1)/2\big)$; wir erhalten dann

$$c_{NS}=\int_0^1\Phi^{-1}\left(\frac{u+1}{2}\right)w_f(u)\,du\ =\int_0^1\left[\Phi^{-1}\left(\frac{u+1}{2}\right)\right]^2 du\ =2\int_0^\infty z^2\varphi(z)\,dz\ =\ 1,$$

und damit ist im Falle einer Normalverteilung $are(NS:t)=1$, da ja $\sigma^2=Var(X_i)=1$ für $f=\varphi$. Bei Vorliegen einer Normalverteilung ist also der Normal-Scores-Test asymptotisch äquivalent zum einfachen t-Test (analog zum Zweistichprobenproblem).

Es gilt $are(NS:t)\geq 1$. Die asymptotische relative Effizienz des Normal-Scores-Tests im Vergleich zum t-Test kann also nie schlechter werden als 1! Dabei gilt das Gleichheitszeichen nur im Falle der Normalverteilung; bei jeder anderen Verteilung ist der Normal-Scores-Test (asymptotisch) besser als der t-Test.

*Begründung:* Mit der Substitution $v=(u+1)/2$ können wir schreiben

$$c_{NS}=\int_0^1\Phi^{-1}\left(\frac{u+1}{2}\right)w_f(u)\,du=2\int_{\frac{1}{2}}^1\Phi^{-1}(v)\ \tilde{w}_f(v)\,dv\quad\text{mit}\quad\tilde{w}_f(v)=-\frac{f'\left[F^{-1}(v)\right]}{f\left[F^{-1}(v)\right]}.$$

Nun sind die beiden Funktionen $\Phi^{-1}(v)$ und $\tilde{w}_f(v)$ antisymmetrisch bezüglich $v=1/2$, und daher gilt

$$c_{NS} = 2\int_{\frac{1}{2}}^{1} \Phi^{-1}(v)\ \tilde{w}_f(v)\,dv = \int_0^1 \Phi^{-1}(v)\ \tilde{w}_f(v)\,dv\,.$$

Die Konstante $c_{NS}$ kann somit dargestellt werden wie im Zweistichprobenfall, und daher gilt wie dort $are(NS:t) \geq 1$.

Im Spezialfall der Normalverteilung ist $f = \varphi$, und $w_f(u) = \Phi^{-1}(u)$, wie wir oben gesehen haben; die Gewichte des Normal-Scores-Tests sind also optimal im Falle einer Normalverteilung.

c) *Vorzeichentest*

Hier ist

$$w(u) = 1 \text{ für } 0 \leq u \leq 1;$$

$$w_i = w\left(\frac{i}{n+1}\right) = 1 \text{ für } i = 1,\ldots,n;$$

Dann gilt

$$L_+ = \sum_{i=1}^{n} V_i \ = \ \#\{i \mid X_i > 0\},$$

und damit ist der lineare Rangtest äquivalent zum Vorzeichentest (vgl. Kapitel 2). Es gilt

$$\tau^2 = \tfrac{1}{4}\int_0^1 w^2(u)\,du = \tfrac{1}{4}.$$

$$c \ = \ c_V \ = \ \frac{I_f}{\tau} \ = \ 4\int_0^{\infty} w'\big(2F(x)-1\big)\,f^2(x)\,dx \ + \ 2\,w(0)\,f(0) \ = \ 2\,f(0)\,.$$

Die asymptotische relative Effizienz des Vorzeichentests im Vergleich zum t-Test beträgt somit

$$are(V:t) = \left(\frac{c_V}{c_t}\right)^2 = 4\sigma^2 f^2(\mu)\,.$$

Es gilt $0 \leq are(V:t) \leq \infty$. Wir haben hier das gleiche Resultat gefunden wie beim Vergleich des einfachen t-Tests mit dem Vorzeichentest im Kapitel 2.

Die zweiseitige Exponentialverteilung hat die Dichte $f(x) = \frac{1}{2}e^{-|x|}$, und es gilt

$$-\frac{f'(x)}{f(x)} = \begin{cases} +1 & \text{für } x > 0 \\ -1 & \text{für } x < 0, \end{cases}$$

und somit gilt

$$w_f(u) \ = \ -\frac{f'\left[F^{-1}\left(\dfrac{u+1}{2}\right)\right]}{f\left[F^{-1}\left(\dfrac{u+1}{2}\right)\right]} \ = \ +1 \text{ für } 0 < u < 1.$$

Dies zeigt, dass die Gewichtung des Vorzeichentests optimal ist bei der zweiseitigen Exponentialverteilung.

d) *Optimale Gewichtung bei einer Cauchy-Verteilung*; *Cauchy-Scores-Test*

Die Dichtefunktion der Cauchy-Verteilung ist gegeben durch

$$f(x)=\frac{1}{\pi(1+x^2)},\ -\infty<x<\infty.$$

Es gilt

$$-\frac{f'(x)}{f(x)}=\frac{2x}{1+x^2}$$

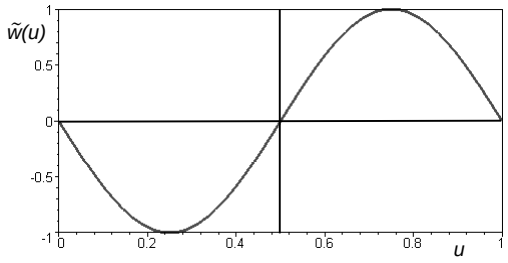
$$F(x)=\frac{1}{\pi}\left(\arctan(x)+\tfrac{1}{2}\right)$$

$$F^{-1}(u)=-\operatorname{cotan}(\pi u)$$

$$\tilde{w}_f(u)=-\frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}=-\sin(2\pi u)$$

$$w_f(u)=\tilde{w}_f\left(\frac{u+1}{2}\right)=-\sin\left[\pi(u+1)\right]=\sin(\pi u).$$

Abb. 1: Gewichtsfunktion $\tilde{w}(u)=-\sin(2\pi u)$




Die optimale Gewichtsfunktion beim Zweistichprobenproblem ist gegeben durch $\tilde{w}(u)=-\sin(2\pi u)$ (vgl. Abb. 1), während die optimale Gewichtsfunktion beim Einstichprobenproblem gegeben ist durch $w(u)=\tilde{w}[(u+1)/2]=\sin(\pi u)$ (rechte Hälfte der Abbildung 1).

Wir betrachten nun den linearen Rangtest mit der Gewichtsfunktion $w(u)=\sin(\pi u)$ ; die zugehörigen Gewichte nennen wir Cauchy-Gewichte (Cauchy-Scores), weil sie optimal sind bei einer Cauchy-Verteilung:

$$w(u)=\sin(\pi u);$$

$$w_i=w\left(\frac{i}{n+1}\right)=\sin\left(\frac{i}{n+1}\pi\right)\ \text{ für }\ i=1,\ldots,n;$$

$$\tau^2=\tfrac{1}{4}\int_0^1 w^2(u)\,du=\tfrac{1}{4}\int_0^1\sin^2(\pi u)\,du=\tfrac{1}{8};$$

$$c=c_{CS}=\frac{I_f}{\tau}=2\sqrt{2}\,I_f=\sqrt{2}\int_0^1 w(u)\,w_f(u)\,du=\sqrt{2}\int_0^1\sin(\pi u)\,w_f(u)\,du.$$

Die asymptotische relative Effizienz des Cauchy-Scores-Tests im Vergleich zum t-Test beträgt

$$are(CS:t)=\left(\frac{c_{CS}}{c_t}\right)^2=\sigma^2c_{CS}^2=2\sigma^2\left(\int_0^1\sin(\pi u)\,w_f(u)\,du\right)^2.$$

Wie im Beispiel b) kann man zeigen, dass die Konstante $c_{CS}$ dargestellt werden kann wie im Zweistichprobenfall.

Unsere Beispiele zeigen, dass die Ergebnisse beim Einstichprobenproblem im wesentlichen äquivalent sind zu den Ergebnissen beim Zweistichprobenproblem, wobei allerdings beim Einstichprobenproblem nur symmetrische Dichtefunktionen zugelassen sind: $f(-x)=f(x)$ für $x\in\mathbb{R}$ .

# Anhang

## A1 Hilfssatz zur stochastischen Konvergenz

Es sei $X_1, X_2, \ldots$ eine Folge von Zufallsgrößen mit $X_n \xrightarrow{P} a$.
Dann existiert eine Folge $\varepsilon_n$ mit $\varepsilon_n \downarrow 0$ so, dass $P\{|X_n - a| > \varepsilon_n\} \to 0$ für $n \to \infty$.

*Beweis:*
Da $X_n \xrightarrow{P} a$, so gilt $P\{|X_n - a| > \varepsilon\} \to 0$ für jedes $\varepsilon > 0$. Es sei $\eta_0 = 1$. Dann gilt offensichtlich $P\{|X_n - a| > \eta_0\} \le \eta_0 = 1$ für $n = 1, 2, \ldots$. Nun sei $k \in \{1, 2, \ldots\}$ und $\eta_k = 2^{-k}$. Dann gilt $P\{|X_n - a| > \eta_k\} \to 0$ für $n \to \infty$ ($\eta_k$ fest), und daher existiert eine natürliche Zahl $n_k$ so, dass $P\{|X_n - a| > \eta_k\} \le \eta_k$ für jedes $n \ge n_k$. Wir setzen $n_0 = 1$ und haben jetzt eine Folge von Paaren $(\eta_k, n_k)$, $k = 0, 1, \ldots$, gefunden mit $\eta_k \downarrow 0$, $n_0 \le n_1 \le \cdots$ und mit $P\{|X_n - a| > \eta_k\} \le \eta_k$ für jedes $n \ge n_k$. Nun setzen wir

$\varepsilon_n = 1 \ (= \eta_0)$ für $n_0 \le n < n_1$
$\varepsilon_n = \frac{1}{2} \ (= \eta_1)$ für $n_1 \le n < n_2$
$\varepsilon_n = \frac{1}{4} \ (= \eta_2)$ für $n_2 \le n < n_3$
…

Dann gilt offensichtlich $\varepsilon_n \downarrow 0$ für $n \to \infty$, und für $n_k \le n < n_{k+1}$ gilt

$$P\{|X_n - a| > \varepsilon_n\} = P\{|X_n - a| > \eta_k\} \le \eta_k \to 0 \text{ für } n \to \infty.$$

## A2 Vier-Felder-Test, Asymptotische Gütefunktion

Es seien $X_1$, $X_2$ unabhängige Zufallsgrößen je mit einer Binomialverteilung $Bi(n, \vartheta_1)$ bzw. $Bi(n, \vartheta_2)$, und weiter sei

(1) $\delta = \sqrt{n}(\vartheta_1 - \vartheta_2)$ und $\bar{\vartheta} = \frac{1}{2}(\vartheta_1 + \vartheta_2)$.

Dann gilt

(2) $\vartheta_1 = \bar{\vartheta} + \dfrac{\delta}{2\sqrt{n}}$ und $\vartheta_2 = \bar{\vartheta} - \dfrac{\delta}{2\sqrt{n}}$.

Für $n \to \infty$ gilt also $\vartheta_1 \to \bar{\vartheta}$ und $\vartheta_2 \to \bar{\vartheta}$. Die Gütefunktion zum Vier-Felder-Test kann geschrieben werden als

Vierfeldertabelle

| | + | – | |
|---|---|---|---|
| $X_1$ | $k_1$ | · | $n$ |
| $X_2$ | $k_2$ | · | $n$ |
| $X$ | $k$ | · | $2n$ |

(3) $$\beta_n(\delta) = P\{\text{Test verwirft } H_0\} = \sum_{k=0}^{n} P\{X=k\} P\{\text{Test verwift } H_0 \,|\, X=k\} = \sum_{k=0}^{n} P\{X=k\} \beta_{n|k}(\delta)$$

(vgl. Abschnitt 6.2). Die Gütefunktion ist abhängig von $n, \delta, \bar{\vartheta}, \alpha$. Wir wollen nun den Grenzwert bestimmen für $n \to \infty$ ($\delta, \bar{\vartheta}, \alpha$ fest). Es gilt

$$\frac{X}{n} = \frac{X_1}{n} + \frac{X_2}{n} \xrightarrow{P} \vartheta_1 + \vartheta_2 = 2\bar{\vartheta} \text{ für } n \to \infty.$$

Mit Hilfe des Satzes in Anhang A1 kann man zeigen (analog wie beim Vorzeichentest, vgl. Abschnitt 4.3), dass der

Grenzwert von $\beta_n(\delta)$ für $n \to \infty$ ($\delta, \bar{\vartheta}, \alpha$ fest)

(4) gleich dem

Grenzwert von $\beta_{n|k}(\delta)$ für $n \to \infty, k/n \to 2\bar{\vartheta}$ ($\delta, \bar{\vartheta}, \alpha$ fest).

Nun ist

$$\beta_{n|k}(\delta) = P\{\text{Test verwirft } H_0 \,|\, X = k\} = P\{X_1 > k_r \,|\, X = k\},$$

und die bedingte Verteilung von $X_1$ gegeben $X = k$ ist die verallgemeinerte hypergeometrische Verteilung $H(2n, n, k, \rho)$, wobei $\rho = \vartheta_1(1-\vartheta_2)/[\vartheta_2(1-\vartheta_1)]$. Leider sind für den Erwartungswert und die Varianz dieser Verteilung keine einfachen Formeln bekannt. Wir können aber die bedingte Gütefunktion $\beta_{n|k}(\delta)$ wie folgt vereinfachen. Es gilt

$$\left.\begin{aligned} X_1 &> k_r \\ X_1 + X_2 &= k \end{aligned}\right\} \Leftrightarrow \begin{cases} 2X_1 > 2k_r \\ X_1 + X_2 = k \end{cases}$$

$$\Leftrightarrow \begin{cases} X_1 - X_2 > 2k_r - k \\ X_1 + X_2 = k \end{cases}$$

Nun sind die Zufallsgrößen $X_1 + X_2$ und $X_1 - X_2$ asymptotische unabhängig (vgl. unten stehenden Hilfssatz), und daher ist der

Grenzwert von $\beta_{n|k}(\delta) = P\{X_1 - X_2 > 2k_r - k \,|\, X_1 + X_2 = k\}$

(5) gleich dem

Grenzwert von $P\{X_1 - X_2 > 2k_r - k\}$.

Es gilt also

(6) $$\lim \beta_n(\delta) = \lim \beta_{n|k}(\delta) = \lim P\{X_1 - X_2 = 2k_r - k\},$$

wobei beim zweiten und dritten Grenzprozess gilt: $n \to \infty, k/n \to 2\bar{\vartheta}$ $(\delta, \bar{\vartheta}, \alpha$ fest$)$.

Nun ist $k_r$ die rechte kritische $\alpha$-Schranke zur hypergeometrischen Verteilung $H(2n, n, k)$. Die standardisierte hypergeometrische Verteilung konvergiert gegen eine standardisierte Normalverteilung, wenn die Varianz gegen unendlich strebt. Wir benötigen daher den Erwartungswert und die Varianz unserer hypergeometrischen Verteilung:

| übliche Notation | beim Vier-Felder-Test |
|---|---|
| $Y \sim H(N, M, n)$ | $Y \sim H(2n, n, k)$ |
| $E(Y) = n\frac{M}{N}$ | $E(Y) = \frac{k}{2} = \mu_0$ |
| $Var(Y) = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$ | $Var(Y) = \frac{k}{4}\frac{2n-k}{2n-1} = \sigma_0^2$ |

Bei unserem Grenzübergang $n \to \infty, k/n \to 2\bar{\vartheta}$ gilt

$$\mu_0 = \frac{k}{2} = n\frac{k}{2n} \sim n\bar{\vartheta} \qquad (\sim: \text{asymptotisch äquivalent})$$

$$\sigma_0^2 = \frac{k}{4}\frac{2n-k}{2n-1} = \frac{n}{2}\frac{k}{2n}\frac{2n-k}{2n-1} \sim \frac{n}{2}\bar{\vartheta}(1-\bar{\vartheta})$$

und damit gilt $\sigma_0^2 \to \infty$; daher haben wir

$$\frac{k_r - \mu_0}{\sigma_0} \to z_{1-\alpha}.$$

Wir schreiben

(7) $$P\{X_1 - X_2 > 2k_r - k\} = P\left\{\frac{X_1 - X_2}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}} > \frac{2k_r - k}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}\right\} = P\{(a) > (b)\}.$$

Nun gilt

$$(a)=\frac{X_1-X_2}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}=\underbrace{\frac{X_1-n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}}}_{\to N(0,1)}\underbrace{\frac{\sqrt{\vartheta_1(1-\vartheta_1)}}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}}_{\to 1}-\underbrace{\frac{X_2-n\vartheta_2}{\sqrt{n\vartheta_2(1-\vartheta_2)}}}_{\to N(0,1)}\underbrace{\frac{\sqrt{\vartheta_2(1-\vartheta_2)}}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}}_{\to 1}+\underbrace{\frac{\sqrt{n}\left(\vartheta_1-\vartheta_2\right)}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}}_{=\frac{\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}},$$

und daher können wir schreiben

$$(a)\ \xrightarrow{d}\ \underbrace{Z_1-Z_2}_{\sim N(0,2)}+\frac{\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}},$$

wobei $Z_1, Z_2$ zwei unabhängige Zufallsgrößen sind je mit einer standardisierten Normalverteilung $N(0,1)$. Weiter gilt

$$(b)=\frac{2k_r-k}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}=2\underbrace{\frac{k_r-\mu_0}{\sigma_0}}_{\to z_{1-\alpha}}\underbrace{\frac{\sigma_0}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}}_{\to 1/\sqrt{2}}+\underbrace{\frac{2\mu_0-k}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}}_{=0},$$

und somit haben wir

$$(b)\ \to\ \sqrt{2}\ z_{1-\alpha}.$$

Aus (7) folgt nun

$$P\left\{X_1-X_2>2k_r-k\right\}\ \to\ P\left\{N(0,2)+\frac{\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}>\sqrt{2}\ z_{1-\alpha}\right\}=P\left\{N(0,1)<-z_{1-\alpha}+\delta\sqrt{\frac{1}{2\bar{\vartheta}(1-\bar{\vartheta})}}\right\},$$

und aufgrund von (6) haben wir damit die gesuchte Grenzfunktion gefunden:

$$(8)\quad \beta_n(\delta)\ \longrightarrow\ P\left\{N(0,1)<-z_{1-\alpha}+\frac{\delta}{\sqrt{2\bar{\vartheta}\left(1-\bar{\vartheta}\right)}}\right\}.$$

Beim allgemeinen Vier-Felder-Test mit beliebigen Stichprobenumfängen $n_1, n_2$ betrachen wir den Grenzübergang

$$n=n_1+n_2\to\infty,\ \frac{n_1}{n}\to\lambda,\ \frac{n_2}{n}\to 1-\lambda,\ \lambda \text{ fest};$$

$$(9)\quad \bar{\vartheta}=\lambda\vartheta_1+(1-\lambda)\vartheta_2 \text{ fest};$$

$$\delta=\sqrt{n}(\vartheta_1-\vartheta_2) \text{ fest}.$$

Auf ähnliche Weise wie beim symmetrischen Fall $n_1=n_2$ kann man zeigen, dass

$$(10)\ \beta_n(\delta)\ \longrightarrow\ P\left\{N(0,1)<-z_{1-\alpha}+\delta\sqrt{\frac{\lambda(1-\lambda)}{\bar{\vartheta}\left(1-\bar{\vartheta}\right)}}\right\}.$$

Man beachte die unterschiedliche Definition von $\delta$ bei (8) und (10):

im ersten Fall: $\delta=\sqrt{n}\left(\vartheta_1-\vartheta_2\right)$ mit $n=n_1=n_2$;

im zweiten Fall: $\delta=\sqrt{n}\left(\vartheta_1-\vartheta_2\right)$ mit $n=n_1+n_2$.

Nun haben wir noch den Hilfssatz zu beweisen, den wir oben bei (5) benutzt haben.

*Hilfssatz*

$X_1$ und $X_2$ seien unabhängige Zufallsgrößen je mit einer $Bi(n,\vartheta_1)$ bzw. $Bi(n,\vartheta_2)$. Dann sind die Zufallsgrößen $X_1+X_2$ und $X_1-X_2$ asymptotische unabhängig.

*Beweis:*

$1^{o}$ $Z_1, Z_2$ seien unabhängige Zufallsgrößen je mit einer standardisierten Normalverteilung $N(0,1)$. Dann sind die beiden Zufallsgrößen $V_1 = Z_1 + Z_2$ und $V_2 = Z_1 - Z_2$ unkorreliert und somit stochastisch unabhängig.

$2^{o}$ $U_1, U_2$ seien unabhängige Zufallsgrößen je mit einer Normalverteilung $N(\mu,\sigma^2)$ bzw. $N(\nu,\sigma^2)$. Dann gilt für die beiden Zufallsgrößen $V_1 = U_1 + U_2$ und $V_2 = U_1 - U_2$

$$E(V_1)E(V_2) = (\mu+\nu)(\mu-\nu) = \mu^2 - \nu^2;$$

$$E(V_1V_2) = E\left(U_1^2\right) - E\left(U_2^2\right) = \sigma^2 + \mu^2 - \left(\sigma^2 + \nu^2\right) = \mu^2 - \nu^2.$$

$V_1$ und $V_2$ sind also unkorreliert und somit stochastisch unabhängig.

$3^{o}$ Es sei

$$U_1 = \frac{X_1 - n\overline{\vartheta}}{\sqrt{n\overline{\vartheta}(1-\overline{\vartheta})}}, \quad U_2 = \frac{X_2 - n\overline{\vartheta}}{\sqrt{n\overline{\vartheta}(1-\overline{\vartheta})}}.$$

Bei unserem Grenzübergang $n \to \infty$ ($\delta, \overline{\vartheta}, \alpha$ fest) gilt gemäß (2)

$$\vartheta_1 = \overline{\vartheta} + \frac{\delta}{2\sqrt{n}} \quad \text{und} \quad \vartheta_2 = \overline{\vartheta} - \frac{\delta}{2\sqrt{n}},$$

und somit erhalten wir

$$U_1 = \frac{X_1 - n\overline{\vartheta}}{\sqrt{n\overline{\vartheta}(1-\overline{\vartheta})}} = \underbrace{\frac{X_1 - n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}}}_{\to N(0,1)} \underbrace{\frac{\sqrt{\vartheta_1(1-\vartheta_1)}}{\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}}_{\to 1} + \underbrace{\frac{\sqrt{n}(\vartheta_1 - \overline{\vartheta})}{\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}}_{= \frac{\delta}{2\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}} \xrightarrow{d} N(\mu,1) \text{ mit } \mu = \frac{\delta}{2\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}.$$

$$U_2 = \frac{X_2 - n\overline{\vartheta}}{\sqrt{n\overline{\vartheta}(1-\overline{\vartheta})}} = \underbrace{\frac{X_2 - n\vartheta_2}{\sqrt{n\vartheta_2(1-\vartheta_2)}}}_{\to N(0,1)} \underbrace{\frac{\sqrt{\vartheta_2(1-\vartheta_2)}}{\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}}_{\to 1} + \underbrace{\frac{\sqrt{n}(\vartheta_2 - \overline{\vartheta})}{\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}}_{= \frac{-\delta}{2\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}} \xrightarrow{d} N(\nu,1) \text{ mit } \nu = \frac{-\delta}{2\sqrt{\overline{\vartheta}(1-\overline{\vartheta})}}.$$

Daher sind die beiden Zufallsgrößen $U_1 + U_2$ und $U_1 - U_2$ aufgrund von $2^{o}$ asymptotisch unabhängig, und somit sind es auch $X_1 + X_2$ und $X_1 - X_2$, was zu zeigen war.

## A3 Zur asymptotischen Gütefunktion beim linearen Rangtest zum Zweistichprobenproblem

Wir wollen hier den Grenzwert von

$$T_2 = \frac{\mu - \mu_0}{\vartheta}$$

bestimmen für

$$N = m + n \;\to\; \infty,$$

(1) $$\frac{m}{N} \;\to\; \lambda,\ \frac{n}{N} \;\to\; 1-\lambda,\ 0<\lambda<1,$$

$$\delta = \vartheta\sqrt{N} \ \text{ fest } \ \left(\Rightarrow \vartheta = \frac{\delta}{\sqrt{N}} \;\to\; 0\right)$$

(vgl. Abschnitt 7.5). Wir bestimmen diesen Grenzwert in zwei Schritten:

(1a) $N \to \infty,\ m/N \to \lambda,\ n/N \to 1-\lambda,\ \vartheta$ fest ;

(1b) $\vartheta = \dfrac{\delta}{\sqrt{N}} \to 0$ bei festem $\delta$ .

Es gilt

(2) $$\mu = \mu(\vartheta) = E_\vartheta(\bar{S}_Y) = E_\vartheta\left[\frac{1}{N}\sum_{j=1}^{n} w\Big(\frac{R_j}{N+1}\Big)\right].$$

Wir wissen bereits, dass beim Grenzübergang (1a) gilt

(3) $$\mu_0 = \mu(0) = E_0(\bar{S}_Y) \;\longrightarrow\; (1-\lambda)\int_0^1 w(u)du \;=\; (1-\lambda)\tilde{w}\,.$$

Nun haben wir den Grenzwert von $\mu = \mu(\vartheta)$ unter der Alternative $\vartheta > 0$ beim Grenzübergang (1a) zu bestimmen. Es sei

$F_m$ die empirische Verteilungsfunktion zu $X_1,\ldots,X_m$

$G_n$ die empirische Verteilungsfunktion zu $Y_1,\ldots,Y_n$ .

Dann gilt

$$R_k = rang(Y_k) \;=\; \#\big\{i \,\big|\, X_i < Y_k\big\} \;+\; \#\big\{j \,\big|\, Y_j \le Y_k\big\} \;=\; mF_m(Y_k) + nG_n(Y_k)\,,$$

und da $n\,dG_n(y) = 1$ ist für $y = Y_j$ , $j = 1,\ldots,n$ , so finden wir die Darstellung

(4)
$$\begin{aligned}\bar{S}_Y \;&=\; \frac{1}{N}\sum_{i=1}^{n} w\Big(\frac{R_j}{N+1}\Big) \;=\; \frac{1}{N}\sum_{i=1}^{n} w\Big[\frac{1}{N+1}\big(mF_m(Y_j) + nG_n(Y_j)\big)\Big] \\ &=\; \frac{n}{N}\int_{-\infty}^{\infty} w\Big[\frac{1}{N+1}\big(mF_m(y) + nG_n(y)\big)\Big]dG_n(y).\end{aligned}$$

Aufgrund des Satzes von Glivenko-Cantelli (Zentralsatz der Statistik) gilt

$$\sup_x \big|F_m(x) - F(x)\big| \;\xrightarrow[m\to\infty]{P}\; 0$$

$$\sup_x \big|G_n(x) - G(x)\big| \;\xrightarrow[n\to\infty]{P}\; 0,$$

und somit ist es plausibel, dass beim Grenzübergang (1a)

$$\bar{S}_Y \;=\; \frac{n}{N}\int_{-\infty}^{\infty} w\Big[\frac{1}{N+1}\big(mF_m(y)+nG_n(y)\big)\Big]dG_n(y)$$
$$\xrightarrow{\;P\;} \;\; (1-\lambda)\int_{-\infty}^{\infty} w\big[\lambda F(y)+(1-\lambda)G(y)\big]dG(y),$$

und dass dann auch der Erwartungswert von $\bar{S}_Y$ gegen diesen Grenzwert konvergiert:

$$\mu=\mu(\vartheta)=E_\vartheta\big(\bar{S}_Y\big) \quad\longrightarrow\quad (1-\lambda)\int_{-\infty}^{\infty} w\big[\lambda F(y)+(1-\lambda)G(y)\big]dG(y) \;=\; \tilde{\mu}(\vartheta)\,.$$

Es ist

$$\tilde{\mu}(0) \;=\; (1-\lambda)\int_{-\infty}^{\infty} w\big[F(y)\big]dF(y) \;=\; (1-\lambda)\int_0^1 w(t)\,dt \;=\; (1-\lambda)\tilde{w} \;=\; \lim\mu(0)\,.$$

Nun gilt also beim Grenzübergang (1a)

$$\mu=\mu(\vartheta)=E_\vartheta(\bar{S}_Y) \longrightarrow (1-\lambda)\int_{-\infty}^{\infty} w\big[\lambda F(y)+(1-\lambda)G(y)\big]dG(y) \;=\; \tilde{\mu}\big(\vartheta\big),$$

$$\mu_0=\mu(0)=E_0(\bar{S}_Y) \longrightarrow (1-\lambda)\tilde{w} \;=\; \tilde{\mu}\big(0\big).$$

$$T_2=\frac{\mu-\mu_0}{\vartheta} \longrightarrow \tilde{T}_2=\frac{\tilde{\mu}(\vartheta)-\tilde{\mu}(0)}{\vartheta}.$$

Wir haben noch den Grenzübergang (1b) durchzuführen, und wir erhalten für $\vartheta\to 0$ offensichtlich

$$\tilde{T}_2=\frac{\tilde{\mu}(\vartheta)-\tilde{\mu}(0)}{\vartheta} \longrightarrow \frac{d}{d\vartheta}\tilde{\mu}(\vartheta)\Big|_{\vartheta=0}=\tilde{\mu}'(0)\,.$$

Damit gilt beim Grenzübergang (1)

$$(5)\quad T_2=\frac{\mu-\mu_0}{\vartheta} \longrightarrow \tilde{\mu}'(0)\,.$$

Nun haben wir noch die Ableitung $\tilde{\mu}'(0)$ zu bestimmen. Es gilt

$$\begin{aligned}\tilde{\mu}\big(\vartheta\big) \;&=\; (1-\lambda)\int_{-\infty}^{\infty} w\big[\lambda F(y)+(1-\lambda)G(y)\big]dG(y)\\ &=\; (1-\lambda)\int_{-\infty}^{\infty} w\big[\lambda F(y)+(1-\lambda)F(y-\vartheta)\big]\,f(y-\vartheta)\,dy\\ &=\; (1-\lambda)\int_{-\infty}^{\infty} w\big[\lambda F(y+\vartheta)+(1-\lambda)F(y)\big]\,f(y)\,dy,\end{aligned}$$

und wir erhalten

$$(6)\quad \tilde{\mu}'\big(0\big) \;=\; \lambda(1-\lambda)\int_{-\infty}^{\infty} w'\big[\lambda F(y)+(1-\lambda)F(y)\big]\,f^2(y)\,dy \;=\; \lambda(1-\lambda)\int_{-\infty}^{\infty} w'\big[F(y)\big]\,f^2(y)\,dy\,.$$

Unter der Voraussetzung $w\big[F(y)\big]\,f(y)\to 0$ für $y\to\pm\infty$ ergibt sich durch partielle Integration

$$(7)\quad \tilde{\mu}'\big(0\big) \;=\; \lambda(1-\lambda)\int_{-\infty}^{\infty} w'\big[F(y)\big]\,f^2(y)\,dy \;=\; \lambda(1-\lambda)\int_{-\infty}^{\infty} w\big[F(y)\big]\,f'(y)\,dy\,,$$

und daraus ergibt sich mit Hilfe der Substitution $F(y)=u$

$$(8)\quad \tilde{\mu}'\big(0\big) \;=\; \lambda(1-\lambda)\int_{-\infty}^{\infty} w'(u)\,f\big[F^{-1}(u)\big]\,du \;=\; \lambda(1-\lambda)\int_{-\infty}^{\infty} w(u)\frac{f'\big[F^{-1}(u)\big]}{f\big[F^{-1}(u)\big]}\,du\,.$$

*Zusammenfassung:*

Falls

$f(x) > 0$ für alle $x \in \mathbb{R}$,

Dichte *f* und Gewichtsfunktion *w* differenzierbar,

$w\big[F(x)\big]\, f(x) \to 0$ für $x \to \pm\infty$,

so gilt

$$T_2 = \frac{\mu - \mu_0}{\vartheta} \longrightarrow \lambda(1-\lambda)\, I_f\,,$$

wobei

$$I_f = \int_{-\infty}^{\infty} w'\big[F(x)\big]\, f^2(x)\, dx = -\int_{-\infty}^{\infty} w\big[F(x)\big]\, f'(x)\, dx$$

$$= \int_0^1 w'(u)\, f\big[F^{-1}(u)\big]\, du = \int_0^1 w(u)\, w_f(u)\, du \quad \text{mit} \quad w_f(u) = -\frac{f'\big[F^{-1}(u)\big]}{f\big[F^{-1}(u)\big]}.$$

## A4 Optimale Gewichtsfunktion

Wir wollen hier zeigen, dass die Konstante *c* bei der asymptotischen Gütefunktion eines linearen Rangtests maximal wird bei der Gewichtsfunktion $w_f$ (vgl. Abschnitt 7.5). Da sich die Konstante *c* nicht ändert bei einer Lineartransformation der Gewichte, so können wir annehmen, dass $\tilde{w} = 0$. Aufgrund der Ungleichung von Cauchy-Schwarz gilt dann für jede beliebige Gewichtsfunktion *w*

$$c = \frac{\int_0^1 w(u)\, w_f(u)\, du}{\left(\int_0^1 w^2(u)\, du\right)^{\frac{1}{2}}} \le \left(\frac{\int_0^1 w^2(u)\, du \int_0^1 w_f^2(u)\, du}{\int_0^1 w^2(u)\, du}\right)^{\frac{1}{2}} = \left(\int_0^1 w_f^2(u)\, du\right)^{\frac{1}{2}}.$$

Für die Gewichtsfunktion $w = w_f$ haben wir

$$\tilde{w}_f = \int_0^1 w_f(u)\, du = -\int_0^1 \frac{f'\big[F^{-1}(u)\big]}{f\big[F^{-1}(u)\big]}\, du = -\int_{-\infty}^{\infty} f'(x)\, dx = 0\,,$$

falls $f(x) \to 0$ für $x \to \pm\infty$. Für $w = w_f$ gilt $c = \left(\int_0^1 w_f^2(u)\, du\right)^{\frac{1}{2}}$, und damit ist für den linearen Rangtest mit der Gewichtsfunktion $w = w_f$ die Konstante *c* maximal.

## A5 Zum Test von Wilcoxon

Die asymptotische relative Effizienz des Tests von Wilcoxon im Vergleich zum doppelten t-Test ist gegeben durch

$$are(W:t) = 12\sigma^2\left(\int f^2(x)dx\right)^2,$$

wobei $\sigma^2 = Var(X_i) = Var(Y_j)$ (vgl. Abschnitt 7.6). Wir wollen hier zeigen, dass für die asymptotische relative Effizienz gilt: $are(W:t) \geq \frac{108}{125} = 0.864$. Wir haben also die folgende Aufgabe zu lösen:

Suche eine Dichte *f* mit

$$\int f^2(x)dx = Minimum$$

unter den Nebenbedingungen

1) $\int f(x)dx = 1$ (*f* soll eine Dichte sein);

2) $\int x^2 f(x)dx = 1.$

Als Lösung dieses Problems werden wir eine symmetrische Dichte finden; die zweite Nebenbedingung besagt dann, dass $\sigma^2 = 1$ ist, und somit ist für diese Dichte $are(W:t)$ minimal.

*Notation:* Im Folgenden schreiben wir kurz $\int f$ anstelle von $\int f(x)dx$ etc.

Wir betrachten den folgenden Ausdruck

$$(*) = \int f^2 + 2b\int x^2 f - 2ba^2\int f\ .$$

Unter den beiden Nebenbedingungen ist $(*) = \int f^2 + 2b(1-a^2)$. Wir suchen nun zunächst bei gegebenem $a, b > 0$ das Minimum von $(*)$ ohne die Nebenbedingungen. Wir können schreiben

$$(*) = \int\left[f^2 + 2b(x^2-a^2)f\right] = \int\limits_{|x|<a}\left[f^2 + 2b(x^2-a^2)f\right] + \int\limits_{|x|\geq a}\left[f^2 + 2b(x^2-a^2)f\right].$$

Das zweite Integral auf der rechten Seite wird minimal falls $f(x) = 0$ für $|x| \geq a$. Das erste Integral kann geschrieben werden als

$$\int\limits_{|x|<a}\left[f^2 + 2b(x^2-a^2)f\right] = \int\limits_{|x|<a}\left[f + b(x^2-a^2)\right]^2 - b^2\int\limits_{|x|<a}\left(x^2-a^2\right)^2.$$

Hier ist das zweite Integral auf der rechten Seite unabhängig von *f*, und der Integrand des ersten Integrals kann nicht negativ werden; dieses Integral wird also minimal, falls der Integrand überall gleich Null ist. Damit haben wir die Dichte gefunden, welche den Ausdruck $(*)$ minimiert:

$$f(x) = \begin{cases} 0 & \text{für } |x| \geq a \\ b(a^2-x^2) & \text{für } |x| < a. \end{cases}$$

Nun wollen wir noch die Konstanten *a, b* so wählen, dass die beiden Nebenbedingungen erfüllt sind:

$$\int f(x)dx = 1 \quad \Rightarrow \quad a^3 b = \tfrac{3}{4}$$
$$\int x^2 f(x)dx = 1 \quad \Rightarrow \quad a^5 b = \tfrac{15}{4},$$

und daraus ergibt sich $a = \sqrt{5}$ und $b = 3\sqrt{5}/100$. Es ist also

$$f(x) = \begin{cases} 0 & \text{für } |x| \geq \sqrt{5} \\ \frac{3\sqrt{5}}{100}(5-x^2) & \text{für } |x| < \sqrt{5}. \end{cases}$$

Diese Dichte ist symmetrisch mit $\mu = 0, \sigma^2 = 1$ und $12\left(\int f^2(x)dx\right)^2 = \frac{108}{125} = 0.864$.

*Bemerkung*

Bei den linearen Rangtests im Kapitel 7 haben wir vorausgesetzt, dass die Dichte *f* überall positiv ist, und dass die Ableitung existiert. Für die oben gefundene Dichte gilt $f(x) = 0$ für $|x| \geq a$ und die Ableitung $f'(a)$ existiert nicht. Diese Dichte ist somit unzulässig für die Untersuchungen im Kapitel 7. Da jedoch das Minimum unseres Extremalproblems in der Familie der zulässigen Dichten nicht kleiner sein kann als in der Familie aller Dichten, so ist die Behauptung $are(W:t) \geq \frac{108}{125} = 0.864$ auch gültig, wenn wir uns auf zulässige Dichten beschränken.

*Maple-Arbeitsblatt* `asrelwil.mws` *zur Berechnung der Konstanten* $a, b$

```
> #
  # Asymptotische relative Effizienz; Minimum beim Test von Wilcoxon
  #
> restart;
> #
  # Minimum von are(W:t) = 0.864
  #
> f := x -> piecewise(abs(x)<a,b*(a^2-x^2));
> f(x);
> #a:=1;
> #b:=2;
> #plot(f(x),x=-3..3);
> Int(x^2*f(x),x=-a..a);
> #simplify(%) assuming a>0,b>0;
> expr1 := value(%) assuming a>0,b>0;
> Int(x*f(x),x=-a..a);
> value(%) assuming a>0,b>0;
> Int(f(x),x=-a..a);
> expr2 := value(%) assuming a>0,b>0;
> eqns := {expr1=1,expr2=1};
> sols := solve(eqns);
> v := allvalues(%);
> v[1];
> v[1,1];
> assign(v[1]);
  a, b;
> a;
> b;
> f(x);
> plot(f(x),x=-3..3);
> int(x^2*f(x),x=-3..3);
> int(f(x)^2,x=-sqrt(5)..sqrt(5));
> INT := int(f(x)^2,x=-3..3);
> # asymptotische relative Effizienz are(t:W)
> 12*INT^2;
> evalf(%);
```

## A6 Zum Normal-Scores-Test

Die asymptotische relative Effizienz des Normal-Scores-Tests im Vergleich zum doppelten t-Test ist gegeben durch

$$are(NS:t) = \sigma^2 c_{NS}^2 ,$$

wobei

$$c_{NS} = I_f = \int_0^1 w(u)\, w_f(u)\, du = \int_{-\infty}^{+\infty} w'\big[F(x)\big]\, f^2(x)\, dx \quad \text{mit} \quad w(u) = \Phi^{-1}(u)$$

(vgl. Abschnitt 7.6). Wir wollen hier zeigen, dass $are(NS:t) \geq 1$; dabei gilt das Gleichheitszeichen nur im Falle der Normalverteilung. Bei jeder anderen Verteilung ist $are(NS:t) > 1$.

Die asymptotische relative Effizienz bleibt invariant bei einer Lineartransformation der Zufallsgößen $X_i$, und daher können wir annehmen, dass $\int x\, f(x)\, dx = 0$ und $\sigma^2 = \int x^2 f(x)\, dx = 1$. Wir haben dann zu zeigen, dass $c_{NS} \geq 1$. Nun ist

$$w'(u) = \frac{1}{\varphi\Big[\Phi^{-1}(u)\Big]},$$

und somit können wir schreiben

$$c_{NS} = \int_{-\infty}^{+\infty} w'\big[F(x)\big]\, f^2(x)\, dx = \int_{-\infty}^{\infty} \frac{f^2(x)}{\varphi\Big[\Phi^{-1}\big(F(x)\big)\Big]}\, dx$$

$$= E\left(\frac{f(X)}{\varphi\Big[\Phi^{-1}\big(F(X)\big)\Big]}\right) = E\left(\frac{1}{\varphi\Big[\Phi^{-1}\big(F(X)\big)\Big] \Big/ f(X)}\right).$$

Aufgrund der Ungleichung von Jensen (siehe unten) gilt für eine positive Zufallsgröße $V$ die Ungleichung $E\big(1/V\big) \geq 1/E(V)$, wobei das Gleichheitszeichen nur gilt, wenn die Zufallsgröße $V$ mit Wahrscheinlichkeit 1 konstant ist. In unserem Fall ist

$$V = \frac{\varphi\Big[\Phi^{-1}\big(F(X)\big)\Big]}{f(X)}.$$

Die Zufallsgröße $V$ kann nur positive Werte annehmen, und im Falle der Normalverteilung ist $f = \varphi$ und $F = \Phi$, und wir erhalten $V \equiv 1$. In diesem Fall gilt also in der Ungleichung von Jensen das Gleichheitszeichen, bei jeder anderen Verteilung jedoch das Größer-Zeichen. Aufgrund der Ungleichung von Jensen gilt also

$$c_{NS} = E\big(1/V\big) \geq \frac{1}{E(V)} \quad \text{d.h.} \quad c_{NS} \geq \frac{1}{E\left(\varphi\Big[\Phi^{-1}\big(F(X)\big)\Big] \Big/ f(X)\right)},$$

oder

$$\frac{1}{c_{NS}} \leq E\left(\frac{\varphi\Big[\Phi^{-1}\big(F(X)\big)\Big]}{f(X)}\right) = \int_{-\infty}^{\infty} \varphi\Big[\Phi^{-1}\big(F(x)\big)\Big]\, dx .$$

Durch partielle Integration erhalten wir mit $v = x$ und

$$u = \varphi\Big[\Phi^{-1}\big(F(x)\big)\Big]$$

$$u' = \frac{\varphi'\Big[\Phi^{-1}\big(F(x)\big)\Big]\, f(x)}{\varphi\Big[\Phi^{-1}\big(F(x)\big)\Big]} = -\Phi^{-1}\big(F(x)\big)\, f(x), \quad \text{da } \varphi'(z) = -z\varphi(z)$$

die Darstellung

$$\int_{-\infty}^{\infty} \varphi\left[\Phi^{-1}(F(x))\right] dx = \int_{-\infty}^{\infty} uv'dx = uv\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} u'v\,dx$$

$$= x\varphi\left[\Phi^{-1}(F(x))\right]\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} x\Phi^{-1}(F(x))\,f(x)dx = A+B.$$

Wir wollen zeigen, dass der Term A verschwindet. Wir können schreiben

$$x\varphi\left[\Phi^{-1}(F(x))\right] = F^{-1}(t)\,\varphi\left(\Phi^{-1}(t)\right) \quad \text{mit der Substitution } F(x)=t$$

$$= F^{-1}(\Phi(w))\,\varphi(w) \quad \text{mit der Substitution } t=\Phi(w),$$

und da das Integral

$$\int_{-\infty}^{\infty} F^{-1}(\Phi(w))\,\varphi(w)\,dw = \int_{0}^{1} F^{-1}(u)\,du = \int_{-\infty}^{\infty} x\,f(x)dx < \infty,$$

so gilt für den Integranden $F^{-1}(\Phi(w))\,\varphi(w) \to 0$ für $w \to \pm\infty$, und damit gilt $A=0$. Somit haben wir

$$\frac{1}{c_{NS}} \le \int_{-\infty}^{\infty} \varphi\left[\Phi^{-1}(F(x))\right] dx = \int_{-\infty}^{\infty} x\Phi^{-1}(F(x))\,f(x)dx\,.$$

Mit Hilfe der Ungleichung von Cauchy-Schwarz erhalten wir

$$\int_{-\infty}^{\infty} x\Phi^{-1}(F(x))\,f(x)dx = \int_{-\infty}^{\infty} x\sqrt{f(x)}\;\Phi^{-1}(F(x))\sqrt{f(x)}\,dx$$

$$\le \left\{\int_{-\infty}^{\infty} x^2 f(x)\,dx \int_{-\infty}^{\infty} \left[\Phi^{-1}(F(x))\right]^2 f(x)\,dx\right\}^{\frac{1}{2}} = 1,$$

denn

$$\int_{-\infty}^{\infty} x^2 f(x)dx = 1$$

$$\int_{-\infty}^{\infty} \left[\Phi^{-1}(F(x))\right]^2 f(x)\,dx = \int_{0}^{1} \left[\Phi^{-1}(u)\right]^2 du = \int_{-\infty}^{\infty} x^2\varphi(x)\,dx = 1\,.$$

In der Ungleichung von Cauchy-Schwarz $(\int uv\,dx)^2 \le \int u^2dx \int v^2dx$ gilt das Gleichheitszeichen nur, falls die beiden Funktionen *u* und *v* identisch sind (bis auf eine Nullmenge). In unserem Falle gilt dies nur, falls $F=\Phi$, d.h. im Falle der Normalverteilung. Somit gilt: Für die Normalverteilung $F=\Phi$ gilt $c_{NS}=1$; für jede andere Verteilung gilt $c_{NS}>1$, und damit ist unsere Behauptung bewiesen.

*Hilfssatz (Ungleichung von Jensen)*

*X* sei eine reelle Zufallsgröße mit $\mu = E(X)$, und $h(x)$ sei eine konvexe Funktion. Dann gilt die Ungleichung $E(h(X)) \ge h(E(X))$.

*Beweis:*

Die Tangente an die Funktion $h(x)$ im Punkte $[\mu, h(\mu)]$ besitze die Gleichung $t(x) = a + bx$. Da die Funktion *h* konvex ist, so gilt $h(x) \ge t(x)$ für jedes *x*, und damit $h(X) \ge t(X) = a + bX$. Daraus folgt: $E(h(X)) \ge a + bE(X) = a + b\mu = h(\mu) = h(E(X))$, was zu beweisen war. Da $h(x) = 1/x$ konvex ist für $x > 0$, so gilt für jede positive Zufallsgröße *X* die Ungleichung $E(1/X) \ge 1/E(X)$. Da $h(x)$ strikt konvex ist, so gilt das Gleichheitszeichen nur, falls $X = const$.

## A7 Zur Gleichverteilung

$U_1,\ldots,U_N$ seien unabhängige Zufallsgrößen je mit einer Gleichverteilung $U(0,1)$ und $U_{(1)},\ldots,U_{(N)}$ die zugehörigen Ordnungsstatistiken. Dann gilt

$$E\left(U_{(i)}\right)=\frac{i}{N+1},\ i=1,\ldots,N$$

(vgl. Abschnitt 7.7).

*Beweis:*

Es sei $u\in(0,1)$ und

$$F_i(u)=P\left\{U_{(i)}<u\right\}\quad\text{und}\quad N_u=\#\left\{j\,\middle|\,U_j<u\right\}.$$

Dann gilt $N_u\sim Bi(N,u)$ und

$$F_i(u)=P\left\{U_{(i)}<u\right\}=P\left\{N_u\geq i\right\}=\sum_{j=i}^{N}p(j)\quad\text{mit}\quad p(j)=\binom{N}{j}u^j(1-u)^{N-j}.$$

Nun gilt

$$\frac{d}{du}p(j)=\binom{N}{j}\left[ju^{j-1}(1-u)^{N-j}-(N-j)u^j(1-u)^{N-j-1}\right]$$

$$=N\left[\binom{N-1}{j-1}u^{j-1}(1-u)^{N-j}-\binom{N-1}{j}u^j(1-u)^{N-j-1}\right]\quad\text{für } j=1,\ldots,N-1;$$

$$\frac{d}{du}p(N)=Nu^{N-1}.$$

Damit folgt

$$f_i(u)=\frac{d}{du}F_i(u)=\sum_{j=i}^{N}\frac{d}{du}p(j)=N\binom{N-1}{i-1}u^{i-1}(1-u)^{N-i}=\frac{N!}{(i-1)!(N-i)!}u^{i-1}(1-u)^{N-i}$$

$$=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}u^{i-1}(1-u)^{N-i}=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}u^{a-1}(1-u)^{b-1}\quad\text{mit } a=i,b=N-i+1.$$

Dies ist die Dichte einer Betaverteilung $Be(a,b)$ mit den Parametern $a=i$ und $b=N-i+1$. Der Erwartungswert dieser Verteilung beträgt

$$E(U_{(i)})=\int_0^1 u\,f_i(u)\,du=\frac{a}{a+b}=\frac{i}{N+1}\quad\text{für } i=1,\ldots,N.$$

## A8 Zur stochastischen Ordnung

Wir wollen hier zeigen, dass die stochastische Ordnung nach Typ I die Ordnung nach Typ II impliziert, d.h. dass

$$X \overset{\text{I}}{\prec} Y \;\Rightarrow\; X \overset{\text{II}}{\prec} Y$$

(vgl. Abschnitt 8.1).

*a) Beweis im stetigen Fall*

$X$ und $Y$ seien unabhängige Zufallsgrößen mit den stetigen Verteilungsfunktionen $F$ und $G$. Dann gilt

$$P\{X = Y\} = 0;$$

$$P\{X < Y\} = \int F(x)\,dG(x);$$

$$P\{X > Y\} = \int G(x)\,dF(x);$$

$$P\{X < Y\} + P\{X > Y\} = \int F(x)\,dG(x) + \int G(x)\,dF(x) = 1.$$

Falls beide Zufallsgrößen $X$ und $Y$ die gleiche Verteilungsfunktion $F$ besitzen, so gilt

$$P\{X < Y\} = P\{X > Y\} = \tfrac{1}{2}, \text{ und somit ist dann } \int F(x)\,dF(x) = \tfrac{1}{2}.$$

Nun sei $X \overset{\text{I}}{\prec} Y$, d.h. $F(x) \geq G(x)\ \forall x$ und $F \not\equiv G$. Dann gilt

$$P\{X < Y\} = \int F(x)\,dG(x) \;\geq\; \int G(x)\,dG(x) \;=\; \tfrac{1}{2};$$

$$P\{X > Y\} = \int G(x)\,dF(x) \;\leq\; \int F(x)\,dF(x) \;=\; \tfrac{1}{2}.$$

Da $F \not\equiv G$, so ist in den vorstehenden Ungleichungen das Gleichheitszeichen nicht möglich (vgl. Hilfssatz); somit ist

$$P\{X < Y\} > \tfrac{1}{2}, \text{ d.h. } P\{X < Y\} > P\{X > Y\},$$

und damit ist $X \overset{\text{II}}{\prec} Y$, was zu beweisen war.

*Hilfssatz:*

$F$ und $G$ seien zwei stetige Verteilungsfunktionen mit $F(x) \geq G(x)\ \forall x$ und mit $F \not\equiv G$. Dann gilt

(1) $\int F(x)\,dG(x) > \tfrac{1}{2}$ d.h. $\int G(x)\,dF(x) < \tfrac{1}{2}$.

*Beweis:*

Da $\int F(x)\,dF(x) = \int G(x)\,dG(x) = \tfrac{1}{2}$, so ist (1) äquivalent zu

(2) $$\int [F(x) - G(x)]\,dG(x) = \int [F(x) - G(x)]\,dF(x) > 0;$$

da die beiden Verteilungsfunktion $F$ und $G$ stetig sind, und da $F \not\equiv G$, so existiert ein Intervall $[a,b]$ mit $F(x) - G(x) > 0$ für $x \in [a,b]$; es kann nun nicht gleichzeitig $F(b) - F(a) = G(b) - G(a) = 0$ sein; somit gilt

$$\int_a^b [F(x) - G(x)]\,dG(x) > 0 \quad \text{oder} \quad \int_a^b [F(x) - G(x)]\,dF(x) > 0.$$

Daraus folgt (2) und schließlich (1), was zu beweisen war.

*b) Beweis im diskreten Fall*

$X$ und $Y$ seien unabhängige Zufallsgrößen mit den möglichen Werten $1, 2, \ldots, r$ und mit

$$P\{X = j\} = p_j \quad \text{und} \quad P\{Y = j\} = q_j .$$

Dann gilt

$$P\{X = Y\} = \sum_{j=1}^{r} p_j q_j ;$$

$$P\{X < Y\} = p_1(q_2 + \cdots + q_r) + p_2(q_3 + \cdots + q_r) + \cdots + p_{r-1} q_r = \sum_{i<j} p_i q_j ;$$

$$P\{X > Y\} = \sum_{i>j} p_i q_j .$$

Offensichtlich ist

$$P\{X < Y\} + P\{X > Y\} + P\{X = Y\} = 1, \text{ und somit gilt } \sum_{i,j=1}^{r} p_i q_j = 1 .$$

Falls beide Zufallsgrößen $X$ und $Y$ die gleiche Verteilungsfunktion $F$ besitzen, so gilt $p_i = q_i$ für $i = 1, \ldots, r$, und wir haben

$$P\{X < Y\} = P\{X > Y\} = \sum_{i<j} p_i p_j .$$

Nun sei $X \overset{\text{I}}{\prec} Y$, d.h. $F(j) \geq G(j)$ für $j = 1, \ldots, r$ und $F \not\equiv G$. Dann gilt

$$P\{X \leq j\} \geq P\{Y \leq j\} \quad \text{für} \quad j = 1, \ldots, r \quad \text{(nicht alle gleich!)}$$

$$\Rightarrow \sum_{j=1}^{r} P\{X \leq j\}\big(P\{X = j\} + P\{Y = j\}\big) > \sum_{j=1}^{r} P\{Y \leq j\}\big(P\{X = j\} + P\{Y = j\}\big)$$

$$\Rightarrow \sum_{i \leq j} p_i p_j + \sum_{i \leq j} p_i q_j > \sum_{i \leq j} q_i p_j + \sum_{i \leq j} q_i q_j \quad (*)$$

Analog haben wir

$$P\{X < j\} \geq P\{Y < j\} \quad \text{für} \quad j = 1, \ldots, r \quad \text{(nicht alle gleich!)}$$

$$\Rightarrow \sum_{j=1}^{r} P\{X < j\}\big(P\{X = j\} + P\{Y = j\}\big) > \sum_{j=1}^{r} P\{Y < j\}\big(P\{X = j\} + P\{Y = j\}\big)$$

$$\Rightarrow \sum_{i<j} p_i p_j + \sum_{i<j} p_i q_j > \sum_{i<j} q_i p_j + \sum_{i<j} q_i q_j \quad (**)$$

Wenn wir nun die beiden Ungleichungen (*) und (**) addieren und dabei berücksichtigen, dass

$$\sum_{i \leq j} p_i p_j + \sum_{i<j} p_i p_j = \sum_{i,j=1}^{r} p_i p_j = 1$$

(das Analoge gilt, wenn wir $p$ durch $q$ ersetzen), so erhalten wir

$$\sum_{i \leq j} p_i q_j + \sum_{i<j} p_i q_j > \sum_{i \leq j} q_i p_j + \sum_{i<j} q_i p_j$$

d.h.

$$2\sum_{i<j} p_i q_j + \sum_{i} p_i q_i > 2\sum_{i<j} q_i p_j + \sum_{i} p_i q_i$$

d.h.

$$2P\{X < Y\} + P\{X = Y\} > 2P\{X > Y\} + P\{X = Y\} ;$$

daraus ergibt sich $P\{X < Y\} > P\{X > Y\}$, und damit ist $X \overset{\text{II}}{\prec} Y$, was zu beweisen war.

*c) Beweis im allgemeinen Fall*

Es seien $F$ und $G$ zwei beliebige Verteilungsfunktionen. Dann gilt

$$P\{X<Y\}=\int F(x-0)\,dG(x);$$

$$P\{X\le Y\}=\int F(x+0)\,dG(x);$$

$$P\{X=Y\}=P\{X\le Y\}-P\{X<Y\}=\int\left[F(x+0)-F(x-0)\right]dG(x);$$

$$P\{X>Y\}=\int G(x-0)\,dF(x).$$

Falls beide Zufallsgrößen $X$ und $Y$ die gleiche Verteilungsfunktion $F$ besitzen, so gilt $F\equiv G$, und wir haben

$$\int F(x-0)\,dF(x)=P\{X<Y\}=P\{X>Y\}$$

(3) $$\int F(x+0)\,dF(x)=P\{X\le Y\}=P\{X=Y\}+P\{X>Y\}$$

$$\int F(x-0)\,dF(x)+\int F(x+0)\,dF(x)=P\{X<Y\}+P\{X=Y\}+P\{X>Y\}=1.$$

Nun sei $X \overset{\mathrm{I}}{\prec} Y$, d.h. $F(x)\ge G(x)\ \forall x$ und $F \not\equiv G$. Dann gilt mit $H(x)=F(x)+G(x)$

$$F(x+0)\ge G(x+0)\ \forall x \text{ und } F\not\equiv G$$

$$\Rightarrow \int F(x+0)\,dH(x) > \int G(x+0)\,dH(x)$$

$$\Rightarrow \int F(x+0)\,dF(x)+\int F(x+0)\,dG(x) > \int G(x+0)\,dF(x)+\int G(x+0)\,dG(x) \quad (*)$$

Analog haben wir

$$F(x-0)\ge G(x-0)\ \forall x \text{ und } F\not\equiv G$$

$$\Rightarrow \int F(x-0)\,dH(x) > \int G(x-0)\,dH(x)$$

$$\Rightarrow \int F(x-0)\,dF(x)+\int F(x-0)\,dG(x) > \int G(x-0)\,dF(x)+\int G(x-0)\,dG(x) \quad (**)$$

Wenn wir nun die beiden Ungleichungen (*) und (**) addieren und dabei berücksichtigen, dass gemäß (3)

$$\int F(x-0)\,dF(x)+\int F(x+0)\,dF(x)=1$$

(das Analoge gilt, wenn wir $F$ durch $G$ ersetzen), so erhalten wir

$$\int F(x+0)\,dG(x)+\int F(x-0)\,dG(x) > \int G(x+0)\,dF(x)+\int G(x-0)\,dF(x)$$

d.h.

$$2P\{X<Y\}+P\{X=Y\} > 2P\{X>Y\}+P\{X=Y\};$$

daraus ergibt sich $P\{X<Y\}>P\{X>Y\}$, und damit ist $X \overset{\mathrm{II}}{\prec} Y$, was zu beweisen war.

## A9 Zum Test von Wilcoxon-Mann-Whitney

Wir wollen hier zeigen, dass zwischen den Testgrößen

$$S_Y = \sum_{j=1}^{n} rang(Y_j) \qquad \text{(mit Mittelbildung bei Bindungen)}$$

und

$$U = N_+ + \tfrac{1}{2} N_0 ,$$

wobei

$$N_+ = \#\left\{(i,j)\middle| X_i < Y_j\right\},$$

$$N_0 = \#\left\{(i,j)\middle| X_i = Y_j\right\},$$

$$N_- = \#\left\{(i,j)\middle| X_i > Y_j\right\},$$

der Zusammenhang

$$S_Y = U + \frac{n(n+1)}{2}$$

besteht (vgl. Abschnitt 8.3).

*Beweis:*
Es seien $x_1,\ldots.x_m, y_1,\ldots, y_n$ die Realisationen von $X_1,\ldots, X_m, Y_1,\ldots, Y_n$. Weiter seien $w_1 < w_2 < \cdots < w_k$ die verschiedenen Werte in der Gesamtstichprobe $x_1,\ldots.x_m, y_1,\ldots, y_n$ und

$$m_j = \#\left\{i\middle| x_i = w_j\right\} = \text{ Anzahl der x-Werte mit } x_i = w_j;$$

$$n_j = \#\left\{i\middle| y_i = w_j\right\} = \text{ Anzahl der y-Werte mit } y_i = w_j .$$

Dann gilt

$$m_j \in \{0,1,\ldots,m\} \quad \text{und} \quad m_1 + \cdots + m_k = m;$$

$$n_j \in \{0,1,\ldots,n\} \quad \text{und} \quad n_1 + \cdots + n_k = n.$$

Wir setzen $b_j = m_j + n_j$ für $j = 1,\ldots,k$; es gilt dann $b_1 + \cdots + b_k = m + n$, und der Vektor $b = (b_1,\ldots,b_k)$ definiert die *Bindungskonfiguration* unserer Gesamtstichprobe $x_1,\ldots.x_m, y_1,\ldots, y_n$.
Nun ist

$$N_0 = \#\left\{(i,j)\middle| x_i = y_j\right\} = m_1 n_1 + m_2 n_2 + \cdots + m_k n_k = \sum_{j=1}^{k} m_j n_j$$

$$N_+ = \#\left\{(i,j)\middle| x_i < y_j\right\} = n_2 m_1 + n_3 (m_1 + m_2) + \cdots + n_k (m_1 + \cdots + m_{k-1}) = \sum_{i<j} m_i n_j$$

und somit ist

$$U = N_+ + \tfrac{1}{2} N_0 = \sum_{i<j} m_i n_j + \tfrac{1}{2} \sum_{j=1}^{k} m_j n_j .$$

Nun wollen wir mit Hilfe der Häufigkeiten $(m_i, n_i)$, $i = 1,\ldots,k$, auch die Testgröße $S_Y$ berechnen mit Mittelbildung bei Bindungen. Zunächst haben wir die gemittelten Rangzahlen zu bestimmen; die nachfolgende Tabelle zeigt, dass

$$rang(w_j) = B_{j-1} + \frac{b_j + 1}{2}, \quad \text{wobei} \quad B_j = b_1 + \cdots + b_j .$$

Daraus ergibt sich

$$S_Y = \sum_{i=1}^{n} rang(y_i) \;=\; n_1\, rang(w_1) + n_2\, rang(w_2) + \cdots + n_k\, rang(w_k)$$
$$= \sum_{j=1}^{k} n_j\, rang(w_j) \;=\; \sum_{j=1}^{k} n_j\, B_{j-1} + \tfrac{1}{2}\sum_{j=1}^{k} n_j\,(b_j+1).$$

Nun ist

$$\sum_{j=1}^{k} n_j\, B_{j-1} = \sum_{j=1}^{k} n_j\,(b_1+\cdots+b_{j-1}) = \sum_{i<j} n_j b_i = \sum_{i<j} n_j\,(m_i+n_i) = \sum_{i<j} m_i n_j + \sum_{i<j} n_i n_j\,,$$
$$\sum_{j=1}^{k} n_j\,(b_j+1) = \sum_{j=1}^{k} n_j\,(m_j+n_j+1) = \sum_{j=1}^{k} m_j n_j + \sum_{j=1}^{k} n_j^2 + n\,,$$

und damit erhalten wir

$$S_Y = \sum_{j=1}^{k} n_j\, B_{j-1} + \tfrac{1}{2}\sum_{j=1}^{k} n_j\,(b_j+1) = \sum_{i<j} m_i n_j + \sum_{i<j} n_i n_j + \tfrac{1}{2}\left[\sum_{j=1}^{k} m_j n_j + \sum_{j=1}^{k} n_j^2 + n\right]$$
$$= \underbrace{\sum_{i<j} m_i n_j + \tfrac{1}{2}\sum_{j=1}^{k} m_j n_j}_{=U} + \tfrac{1}{2}\Bigg[\underbrace{2\sum_{i<j} n_i n_j + \sum_{j=1}^{k} n_j^2}_{=n^2}\Bigg] + \tfrac{1}{2}n \;=\; U + \tfrac{1}{2}n(n+1),$$

was zu zeigen war.

**Tabelle** zur Berechnung der durchschnittlichen Rangzahlen

| | |
|---|---|
| $w_1$ | $b_1 = m_1 + n_1$ Stichprobenwerte besitzen den Wert $w_1$;<br>Rangzahlen $1, 2, \ldots, b_1$ fallen auf den Wert $w_1$;<br>$rang(w_1) = \dfrac{1+2+\cdots+b_1}{b_1} = \dfrac{b_1+1}{2}$; |
| $w_2$ | $b_2 = m_2 + n_2$ Stichprobenwerte besitzen den Wert $w_2$;<br>Rangzahlen $b_1+1, b_1+2, \ldots, b_1+b_2$ fallen auf den Wert $w_2$;<br>$rang(w_2) = \dfrac{(b_1+1)+(b_1+2)+\cdots+(b_1+b_2)}{b_2} = b_1 + \dfrac{b_2+1}{2}$; |
| $\vdots$ | $\vdots$ |
| $w_j$ | $b_j = m_j + n_j$ Stichprobenwerte besitzen den Wert $w_j$;<br>Rangzahlen $B_{j-1}+1, B_{j-1}+2, \ldots, B_{j-1}+b_j$ fallen auf den Wert $w_j$ mit $B_j = b_1+\cdots+b_j$;<br>$rang(w_j) = \dfrac{(B_{j-1}+1)+(B_{j-1}+2)+\cdots+(B_{j-1}+b_j)}{b_j} = B_{j-1} + \dfrac{b_j+1}{2}$; |
| $\vdots$ | $\vdots$ |
| $w_k$ | $b_k = m_k + n_k$ Stichprobenwerte besitzen den Wert $w_k$;<br>Rangzahlen $B_{k-1}+1, B_{k-1}+2, \ldots, B_{k-1}+b_k$ fallen auf den Wert $w_k$;<br>$rang(w_k) = \dfrac{(B_{k-1}+1)+(B_{k-1}+2)+\cdots+(B_{k-1}+b_k)}{b_k} = B_{k-1} + \dfrac{b_k+1}{2}$. |

## A10 Zur asymptotischen Gütefunktion beim linearen Rangtest zum Einstichprobenproblem

Wir wollen hier den Grenzwert von

$$T_2 = \frac{\mu - \mu_0}{\vartheta}$$

bestimmen für

$n \to \infty$,

(1) $\delta = \vartheta\sqrt{n}$ fest $\left(\Rightarrow \vartheta = \frac{\delta}{\sqrt{n}} \to 0\right)$

(vgl. Abschnitt 9.5). Wir bestimmen diesen Grenzwert in zwei Schritten:

(1a) $n \to \infty$, $\vartheta$ fest ;

(1b) $\vartheta = \frac{\delta}{\sqrt{n}} \to 0$ bei festem $\delta$ .

Es gilt

$$(2) \quad \mu = \mu(\vartheta) = E_\vartheta(\bar{L}_+) = E_\vartheta\left[\frac{1}{n}\sum_{i=1}^{n} V_i\, w\left(\frac{R_i}{n+1}\right)\right].$$

Wir wissen bereits, dass beim Grenzübergang (1a) gilt

$$(3) \quad \mu_0 = \mu(0) = E_0(\bar{L}_+) \longrightarrow \tfrac{1}{2}\int_0^1 w(u)du = \tilde{w}.$$

Nun haben wir den Grenzwert von $\mu = \mu(\vartheta)$ unter der Alternative $\vartheta > 0$ beim Grenzübergang (1a) zu bestimmen. Es seien

$X_1, \ldots, X_n$ unabhängige Zufallsgrößen je mit der Verteilungsfunktion $G(x) = F(x - \vartheta)$ ;

$x_1, \ldots, x_n$ die zugehörigen Realisationen;

$G_n(x) =$ empirische Verteilungsfunktion zu $x_1, \ldots, x_n = \frac{1}{n} \times \#\{i \,|\, x_i \le x\}$

$H(x) =$ Verteilungsfunktion zu $|X_i|$

$= P\{|X_i| \le x\} = P\{-x \le X_i \le x\} = G(x) - G(-x),\ x \ge 0$

$H_n(x) =$ empirische Verteilungsfunktion zu $|x_1|, \ldots, |x_n| = \frac{1}{n} \times \#\{i \,|\, |x_i| \le x\}$.

Dann gilt für $x_j > 0$

$$n H_n(x_j) = \#\{i \,|\, |x_i| \le x_j\} = r_j = rang(|x_j|).$$

Somit ist

$$\bar{\ell}_+ = \frac{1}{n}\sum_{j=1}^{n} v_j\, w\left(\frac{r_j}{n+1}\right) = \frac{1}{n}\sum_{j \in M_+} w\left(\frac{r_j}{n+1}\right) \quad \text{wobei} \quad M_+ = \{j \,|\, x_j > 0\}$$

$$= \frac{1}{n}\sum_{j \in M_+} w\left[\frac{n}{n+1} H_n(x_j)\right]$$

$$= \int_0^\infty w\left[\frac{n}{n+1} H_n(x)\right] dG_n(x) \quad \text{denn } dG_n(x) = \frac{1}{n} \text{ für } x = x_j .$$

Da dies für beliebige Stichprobenwerte gilt, so haben wir die Darstellung

$$\bar{L}_+ = \int_0^\infty w\Big[\frac{n}{n+1} H_n(x)\Big]\, dG_n(x)\,,$$

wobei

$$G_n(x) = \text{ empirische Verteilungsfunktion zu } X_1,\ldots,X_n \;=\; \frac{1}{n}\times \#\big\{i \,\big|\, X_i \le x\big\}$$

$$H_n(x) = \text{ empirische Verteilungsfunktion zu } |X_1|,\ldots,|X_n| \;=\; \frac{1}{n}\times \#\big\{i \,\big|\, |X_i| \le x\big\}\,.$$

Aufgrund des Satzes von Glivenko-Cantelli (Zentralsatz der Statistik) gilt für $n \to \infty$

$$\sup_x \big|G_n(x) - G(x)\big| \;\xrightarrow{\;P\;}\; 0$$

$$\sup_x \big|H_n(x) - H(x)\big| \;\xrightarrow{\;P\;}\; 0,$$

und somit ist es plausibel, dass für $n \to \infty$

$$\bar{L}_+ = \int_0^\infty w\Big[\frac{n}{n+1} H_n(x)\Big]\, dG_n(x) \;\xrightarrow{\;P\;}\; \int_0^\infty w\big[H(x)\big]\, dG(x)\,,$$

und dass dann auch der Erwartungswert von $\bar{L}_+$ gegen diesen Grenzwert konvergiert:

$$\mu = \mu(\vartheta) = E_\vartheta(\bar{L}_+) \;\longrightarrow\; \int_0^\infty w\big[H(x)\big]\, dG(x) = \int_0^\infty w\big[G(x) - G(-x)\big]\, dG(x) = \tilde{\mu}(\vartheta)\,.$$

Es ist

$$\tilde{\mu}(0) = \int_0^\infty w\big(F(x) - F(-x)\big)\, dF(x) = \int_0^\infty w\big(2F(x) - 1\big)\, dF(x) = \tfrac{1}{2}\int_0^1 w(u)\, du = \tilde{w} = \lim_{n\to\infty} \mu(0)\,,$$

und somit haben wir beim Grenzübergang (1a) und (1b)

$$T_2 = \frac{\mu(\vartheta) - \mu(0)}{\vartheta} \;\xrightarrow[n\to\infty]{}\; \frac{\tilde{\mu}(\vartheta) - \tilde{\mu}(0)}{\vartheta} \;\xrightarrow[\vartheta\to 0]{}\; \tilde{\mu}'(0)\,.$$

Damit haben wir den gesuchten Grenzwert gefunden. Wir wollen jetzt diesen Grenzwert ausführlicher darstellen. Es gilt

$$\begin{aligned}
\tilde{\mu}(\vartheta) &= \int_0^\infty w\big[G(x) - G(-x)\big]\, dG(x) \\
&= \int_0^\infty w\big[F(x-\vartheta) - F(-x-\vartheta)\big]\, f(x-\vartheta)\, dx \\
&= \int_{-\vartheta}^\infty w\big[F(x) - F(-x-2\vartheta)\big]\, f(x)\, dx \quad \text{(Substitution: } x - \vartheta \;\to\; x\text{)},
\end{aligned}$$

und wir erhalten

$$\begin{aligned}
\tilde{\mu}'(0) &= \frac{d}{d\vartheta}\tilde{\mu}(\vartheta)\Big|_{\vartheta=0} = 2\int_0^\infty w'\big[F(x) - F(-x)\big]\, f^2(x)\, dx \;+\; w(0)\, f(0) \\
&= 2\int_0^\infty w'\big[2F(x) - 1\big]\, f^2(x)\, dx \;+\; w(0)\, f(0).
\end{aligned}$$

*Zusammenfassung:*

Falls

$f(x) > 0$ für alle $x \in \mathbb{R}$,

Dichte $f$ und Gewichtsfunktion $w$ differenzierbar,

$w\big[2F(x)-1\big]\, f(x) \to 0$ für $x \to \infty$,

so gilt

$$T_2 = \frac{\mu - \mu_0}{\vartheta} \longrightarrow \tilde{\mu}'(0),$$

wobei

$$\tilde{\mu}'(0) = 2\int_0^\infty w'\big[2F(x)-1\big]\, f^2(x)\, dx + w(0)\, f(0) = -\int_0^\infty w\big[2F(x)-1\big]\, f'(x)\, dx$$

$$= \int_0^1 w'(u)\, f\left[F^{-1}\left(\frac{u+1}{2}\right)\right] du + w(0)\, f(0) = \tfrac{1}{2}\int_0^1 w(u)\, w_f(u)\, du$$

wobei

$$w_f(u) = -\frac{f'\left[F^{-1}\left(\frac{u+1}{2}\right)\right]}{f\left[F^{-1}\left(\frac{u+1}{2}\right)\right]}.$$

## A11 Optimale Gewichtsfunktion beim Einstichprobenproblem

Wir wollen hier zeigen, dass die Konstante $c$ bei der asymptotischen Gütefunktion eines linearen Rangtests beim Einstichprobenproblem maximal wird bei der Gewichtsfunktion $w_f$ (vgl. Abschnitt 9.5). Aufgrund der Ungleichung von Cauchy-Schwarz gilt für jede beliebige Gewichtsfunktion $w$

$$c = \frac{\int_0^1 w(u)\, w_f(u)\, du}{\left(\int_0^1 w^2(u)\, du\right)^{\frac{1}{2}}} \le \left(\frac{\int_0^1 w^2(u)\, du \int_0^1 w_f^2(u)\, du}{\int_0^1 w^2(u)\, du}\right)^{\frac{1}{2}} = \left(\int_0^1 w_f^2(u)\, du\right)^{\frac{1}{2}}.$$

Für die Gewichtsfunktion $w = w_f$ haben wir

$$c = \left(\int_0^1 w_f^2(u)\, du\right)^{\frac{1}{2}},$$

und damit ist für den linearen Rangtest mit der Gewichtsfunktion $w = w_f$ die Konstante $c$ maximal.

# Literatur

Büning, H. (1991). Robuste und adaptive Tests. Walter de Gruyter, Berlin

Büning, H. und Trenkler, G. (1994). Nichtparametrische statistische Methoden, 2. Auflage. Walter de Gruyter, Berlin.

Hettmansperger, T.P. (1984). Statistical Inference Based on Ranks. Wiley, New York.

Lee, A.J. (1990). U-Statistics. Theory and Practice. Marcel Dekker, New York.

Lehmann, E.L. (1975). Nonparametrics: Statistical Methods Based on Ranks. Holden-Day, California.

Schaich, E. und Hamerle, A. (1084). Verteilungsfreie statistische Prüfverfahren. Springer, Berlin.