

Vorlesung Test- und Schätztheorie II, SS 2002

**Leo Knüsel, Universität München**

## Inhalt

<b>1. Poissonverteilung: Zwei-Stichprobenproblem .....</b>	<b>1</b>
1. Problemstellung .....	1
2. Lösung zum Testproblem .....	1
3. Gütefunktion des Tests .....	2
4. Zur Optimalität des Tests bei den einseitigen Fragestellungen 1a) und 1b) .....	2
5. Verallgemeinerung .....	2
<b>2. Poissonverteilung: Prognoseproblem.....</b>	<b>4</b>
1. Problemstellung .....	4
2. Lösung zum Prognoseproblem (duale Form des Zwei-Stichprobenproblems).....	4
3. Verallgemeinerung .....	5
4. Näherungsformeln .....	5
<b>3. Binomialverteilung: Zweistichprobenproblem und Prognoseproblem .....</b>	<b>6</b>
1. Problemstellung .....	6
2. Lösung zum Testproblem .....	6
3. Näherungsverfahren.....	7
4. Zur Optimalität des Tests bei den einseitigen Fragestellungen 1a) und 1b) .....	7
5. Prognoseproblem .....	7
6. Näherungsformeln für die Prognosegrenzen .....	8
7. Vertrauensgrenzen für den Parameter $\rho$ (= Odds Ratio) .....	8
8. Verallgemeinerung .....	8
<b>4. Binomialverteilung: Asymptotische Gütefunktion zum Vier-Felder-Test.....</b>	<b>9</b>
1. Testproblem.....	9
2. Lösung zum Testproblem .....	9
3. Asymptotische Gütefunktion im Spezialfall $n_1 = n_2$ .....	9
4. Zur Unabhängigkeit von $X_1 + X_2$ und $X_1 - X_2$ .....	11
5. Asymptotische Gütefunktion, Fortsetzung .....	12
6. Asymptotische Gütefunktion im allgemeinen Fall .....	14
7. Notwendiger Stichprobenumfang .....	18
<b>5. Normalverteilung: Vergleich von zwei Mittelwerten .....</b>	<b>20</b>
1. Problemstellung .....	20
2. Lösung .....	20
a) Varianzen bekannt .....	20
b) Varianzen unbekannt, jedoch beide gleich .....	21
c) Varianzen unbekannt (Behrens-Fisher-Problem) .....	21
3. Optimale Aufteilung des Stichprobenumfangs bei gegebenen Varianzen .....	22
<b>6. Normalverteilung: Vergleich von k Mittelwerten; einfache Varianzanalyse (ANOVA).....</b>	<b>23</b>
1. Problemstellung .....	23
2. Lösung .....	23
3. Spezialfall $k=2$ (Zwei-Stichprobenproblem).....	24
<b>7. Normalverteilung: Vergleich von zwei Varianzen .....</b>	<b>26</b>
1. Problemstellung .....	26
2. Lösung .....	26
a) Mittelwerte bekannt .....	26
b) Mittelwerte unbekannt.....	27

<b>8. Multinomialverteilung: Einstichprobenproblem .....</b>	<b>28</b>
1. Zusammenhang zwischen Multinomial- und Chi-Quadrat-Verteilung.....	28
2. Einstichprobenproblem bei der Multinomialverteilung.....	28
3. Klassischer Chi-Quadrat-Test.....	28
4. Exakter Chi-Quadrat-Test.....	29
5. Weitere Testkriterien .....	29
6. Monte-Carlo-Variante .....	30
7. Poissonmodell.....	30
8. Chi-Quadrat-Anpassungstest (Goodness of fit test) .....	30
<b>9. Multinomialverteilung: Mehrstichprobenproblem .....</b>	<b>31</b>
1. Problemstellung.....	31
2. Klassischer Chi-Quadrat-Test.....	31
3. Exakter Chi-Quadrat-Test.....	31
4. Einfaches Modell für die verallgemeinerte hypergeometrische Verteilung.....	32
5. Weitere Testkriterien .....	33
6. Monte-Carlo-Variante .....	33
7. Kontingenzmodell .....	34
8. Poissonmodell.....	35
<b>10. Vorzeichentest und Test von McNemar .....</b>	<b>36</b>
1. Vorzeichentest .....	36
2. Test von McNemar (Symmetriepfung in einer Vier-Felder-Tafel): .....	36
<b>11. Asymptotische relative Effizienz zwischen t-Test und Vorzeichentest .....</b>	<b>38</b>
1. t-Test und Vorzeichentest.....	38
2. Asymptotische Gütefunktion des t-Tests.....	38
3. Asymptotische Gütefunktion des Vorzeichentests .....	39
4. Asymptotische relative Effizienz.....	39
<b>12. Unabhängige oder verbundene Stichproben?</b>	
<b>Vier-Felder-Test oder Test von McNemar?.....</b>	<b>41</b>
1. Problemstellung.....	41
2. Versuchsplan mit unabhängigen Stichproben; Vier-Felder-Test.....	41
3. Versuchsplan mit verbundenen Stichproben; Test von McNemar .....	42
4. Asymptotische relative Effizienz.....	43
<b>13. Zur Optimalität von erwartungstreuen Schätzfunktionen .....</b>	<b>45</b>
1. Ungleichung von Cramér-Rao.....	45
2. Satz von Rao.....	46
3. Satz von Rao-Blackwell .....	46
4. Satz von Lehmann-Scheffé.....	46
5. Zum durchschnittlichen quadratischen Fehler.....	47
<b>14. Informationsmatrix bei mehrdimensionalem Parameter .....</b>	<b>48</b>
1. Ausgangssituation.....	48
2. Mathematische Grundlagen .....	48
3. (Fisher-) Informationsmatrix .....	49
4. Cramér-Rao-Ungleichung im mehrdimensionalen Fall.....	49
5. Asymptotische Verteilung von Maximum-Likelihood-Schätzungen .....	50
<b>15. Zur empirischen Verteilungsfunktion .....</b>	<b>51</b>
1. Einführung.....	51
2. Empirische Verteilungsfunktion.....	51
3. Satz von Glivenko-Cantelli (Hauptsatz der Statistik).....	51
4. Satz von Kolmogoroff .....	52
5. Test von Kolmogoroff (Einstichprobenproblem) .....	53
6. Satz von Kolmogoroff-Smirnoff.....	54
7. Test von Kolmogoroff-Smirnoff (Zweistichprobenproblem).....	55

<b>16. Über Dichteschätzung .....</b>	<b>56</b>
1. Einführung.....	56
2. Histogramm-Schätzer .....	56
3. Kern-Dichteschätzer .....	57
<b>17. Invarianzprinzipien, Pitman-Schätzer .....</b>	<b>58</b>
1. Lagemodelle .....	58
2. Äquivalenzklassen, Translationsbahnen .....	58
3. Lösung des Optimalitätsproblems, Teil 1 .....	59
4. Lösung der Optimalitätsaufgabe, Teil 2 .....	62
5. Beispiele .....	64
<b>18. Zur Bayes'schen Statistik.....</b>	<b>69</b>
1. Binomialverteilung mit Beta-Verteilung als a-priori-Verteilung.....	69
1.1 Ausgangssituation .....	69
1.2 A-posteriori-Verteilung von $\vartheta_{rv}$ .....	69
1.3 Randverteilung von $X$ ; Rechnen mit bedingten Erwartungswerten.....	71
1.4 Optimaler Bayes-Schätzer bei quadratischer Verlustfunktion .....	73
2. Poissonverteilung mit Gamma-Verteilung als a-priori-Verteilung.....	74
2.1 Ausgangssituation .....	74
2.2 A-posteriori-Verteilung von $\lambda_{rv}$ .....	74
2.3 Nicht-informative a-priori-Verteilung.....	76
2.4 Randverteilung von $X$ ; Rechnen mit bedingten Erwartungswerten.....	77
2.5 Optimaler Bayes-Schätzer bei quadratischer Verlustfunktion .....	78
3. Multinomialverteilung mit Dirichlet-Verteilung als a-priori-Verteilung.....	79
3.1 Dirichlet-Verteilung .....	79
3.2 Bayes-Modell .....	80
<b>19. Stochastische Ordnung, Test von Wilcoxon.....</b>	<b>83</b>
1. Stochastische Ordnung .....	83
2. Test von Wilcoxon (-Mann-Whitney); Problemstellung .....	87
3. Test von Wilcoxon im stetigen Fall .....	87
4. Test von Wilcoxon im allgemeinen Fall.....	87

# Vorlesung Test- und Schätztheorie II, SS 2002

Leo Knüsel, Universität München

## 1. Poissonverteilung: Zwei-Stichprobenproblem

### 1. Problemstellung

$$X_1 \sim \text{Po}(\lambda_1)$$

$$X_2 \sim \text{Po}(\lambda_2)$$

$X_1, X_2$  unabhängig

$\lambda_1, \lambda_2$  unbekannt

1a)  $H_0: \lambda_1 \leq \lambda_2$  gegen  $H_1: \lambda_1 > \lambda_2$

1b)  $H_0: \lambda_1 \geq \lambda_2$  gegen  $H_1: \lambda_1 < \lambda_2$

1c)  $H_0: \lambda_1 = \lambda_2$  gegen  $H_1: \lambda_1 \neq \lambda_2$

$k_1, k_2$  Realisationen von  $X_1, X_2$

Entscheide aufgrund der Realisationen  $(k_1, k_2)$ , ob  $H_0$  oder  $H_1$  vorliegt.

### 2. Lösung zum Testproblem

Grundidee des Tests:

$$X_1 | X = k \sim \text{Bi}(k, \vartheta) \text{ wobei } X = X_1 + X_2, k = k_1 + k_2 \text{ und } \vartheta = \lambda_1 / (\lambda_1 + \lambda_2)$$

$$\lambda_1 \leq \lambda_2 \Leftrightarrow \vartheta \leq \frac{1}{2}$$

$$\lambda_1 > \lambda_2 \Leftrightarrow \vartheta > \frac{1}{2}$$

Testvorschrift zu vorgegebenem Signifikanzniveau  $\alpha$ :

Berechne  $k = k_1 + k_2$

Bestimme rechte und linke kritische  $\alpha$ -Schranken zu  $\text{Bi}(k, \frac{1}{2})$ :

$$(k_r, \delta_r) \text{ so: } \mathbb{P}\left\{\text{Bi}\left(k, \frac{1}{2}\right) > k_r\right\} + \delta_r \mathbb{P}\left\{\text{Bi}\left(k, \frac{1}{2}\right) = k_r\right\} = \alpha$$

$$(k_\ell, \delta_\ell) \text{ so: } \mathbb{P}\left\{\text{Bi}\left(k, \frac{1}{2}\right) < k_\ell\right\} + \delta_\ell \mathbb{P}\left\{\text{Bi}\left(k, \frac{1}{2}\right) = k_\ell\right\} = \alpha.$$

Aus Symmetriegründen gilt:  $k_\ell + k_r = k$  und  $\delta_\ell = \delta_r$ .

Testfunktion zum Testproblem 1a):

$$\tau(k_1, k_2) = \begin{cases} 1 & \text{falls } k_1 > k_r \\ \delta_r & \text{falls } k_1 = k_r \\ 0 & \text{falls } k_1 < k_r \end{cases}$$

Für  $k_1 = k_2 = k = 0$ :  $\tau(k_1, k_2) \equiv \alpha$ .

Analog für die Probleme 1b) und 1c).

$p$ -Werte (für nichtrandomisierte Testversion):

$$\text{rechter } p\text{-Wert: } p_r = P\left(\text{Bi}\left(k, \frac{1}{2}\right) \geq k_1\right)$$

$$\text{linker } p\text{-Wert: } p_\ell = P\left(\text{Bi}\left(k, \frac{1}{2}\right) \leq k_1\right).$$

### 3. Gütefunktion des Tests

$$\begin{aligned} \beta(\lambda_1, \lambda_2) &= P_{(\lambda_1, \lambda_2)} \{ \text{Test verwirft } H_0 \} \\ &= \sum_{k=0}^{\infty} P_{(\lambda_1, \lambda_2)} \{ X=k \} P_{(\lambda_1, \lambda_2)} \{ \text{Test verwirft } H_0 \mid X=k \} \end{aligned}$$

Nun gilt:

$$X = X_1 + X_2 \sim \text{Po}(\lambda) \quad \text{mit} \quad \lambda = \lambda_1 + \lambda_2$$

$$\begin{aligned} P_{(\lambda_1, \lambda_2)} \{ \text{Test verwirft } H_0 \mid X=k \} &= \text{Gütefunktion des bedingten Tests gegeben } k = k_1 + k_2 \\ &= \beta_k(\vartheta) \quad (\text{nur abhängig von } k \text{ und } \vartheta = \lambda_1 / (\lambda_1 + \lambda_2)). \end{aligned}$$

Daher kann die Gütefunktion des globalen Tests geschrieben werden als

$$\beta(\lambda_1, \lambda_2) = \sum_{k=0}^{\infty} P_\lambda \{ X=k \} \beta_k(\vartheta).$$

### 4. Zur Optimalität des Tests bei den einseitigen Fragestellungen 1a) und 1b)

- 1) Der Test ist optimal (UMP- $\alpha$ -Test, uniformly most powerful) auf jeder Schicht
 
$$S_k = \{(k_1, k_2) \mid k_1 + k_2 = k\}.$$
- 2) Der Test ist ähnlich auf der Grenze (similar on the boundary):
 
$$\beta(\lambda_1, \lambda_2) \equiv \alpha \quad \text{falls } \lambda_1 = \lambda_2.$$
- 3) Der Test ist unverfälscht (unbiased):
 
$$\begin{aligned} \beta(\lambda_1, \lambda_2) &\leq \alpha \quad \text{falls } H_0 \text{ wahr;} \\ \beta(\lambda_1, \lambda_2) &> \alpha \quad \text{falls } H_1 \text{ wahr.} \end{aligned}$$
- 4) Der Test besitzt die größtmögliche Macht unter allen unverfälschten Tests (UMPU- $\alpha$ -Test, uniformly most powerful unbiased).

### 5. Verallgemeinerung

$$X_1 \sim \text{Po}(\lambda_1)$$

$$X_2 \sim \text{Po}(\lambda_2)$$

$X_1, X_2$  unabhängig

$\lambda_1, \lambda_2$  unbekannt

1a)  $H_0: \lambda_1 \leq c\lambda_2$  gegen  $H_1: \lambda_1 > c\lambda_2$  (Konstante  $c$  bekannt)

1b)  $H_0: \lambda_1 \geq c\lambda_2$  gegen  $H_1: \lambda_1 < c\lambda_2$  (Konstante  $c$  bekannt)

1c)  $H_0: \lambda_1 = c\lambda_2$  gegen  $H_1: \lambda_1 \neq c\lambda_2$  (Konstante  $c$  bekannt)

$k_1, k_2$  Realisationen von  $X_1, X_2$

Entscheide aufgrund der Realisationen  $(k_1, k_2)$ , ob  $H_0$  oder  $H_1$  vorliegt.

Grundidee des Tests:

$$X_1 | X = k \sim \text{Bi}(k, \vartheta) \text{ wobei } X = X_1 + X_2, k = k_1 + k_2 \text{ und } \vartheta = \lambda_1 / (\lambda_1 + \lambda_2)$$

$$\lambda_1 \leq c\lambda_2 \Leftrightarrow \vartheta \leq \vartheta_0 = \frac{c}{1+c} \text{ (gegeben!)}$$

$$\lambda_1 > c\lambda_2 \Leftrightarrow \vartheta > \vartheta_0 = \frac{c}{1+c}$$

Testvorschrift zu vorgegebenem Signifikanzniveau  $\alpha$ :

Berechne  $k = k_1 + k_2$

Bestimme rechte und linke kritische  $\alpha$ -Schranken zu  $\text{Bi}(k, \vartheta_0)$  mit  $\vartheta_0 = \frac{c}{1+c}$ ;

weiter wie im symmetrischen Fall  $c = 1$  d.h.  $\vartheta_0 = \frac{1}{2}$ .

$p$ -Werte (für nichtrandomisierte Testversion):

$$\text{rechter } p\text{-Wert: } p_r = \text{P}(\text{Bi}(k, \vartheta_0) \geq k_1) \text{ mit } \vartheta_0 = \frac{c}{1+c}$$

$$\text{linker } p\text{-Wert: } p_\ell = \text{P}(\text{Bi}(k, \vartheta_0) \leq k_1) \text{ mit } \vartheta_0 = \frac{c}{1+c}.$$

## 2. Poissonverteilung: Prognoseproblem

### 1. Problemstellung

$$X_1 \sim \text{Po}(\lambda)$$

$$X_2 \sim \text{Po}(\lambda)$$

$X_1, X_2$  unabhängig

$\lambda$  unbekannt

$k_1$  Realisation von  $X_1$  ( $k_1$  bekannt)

$k_2$  Realisation von  $X_2$  ( $k_2$  noch nicht bekannt)

Bestimme aufgrund der Realisation  $k_1$  Prognosegrenzen für  $X_2$  mit einem vorgegebenen Sicherheitsgrad.

### 2. Lösung zum Prognoseproblem (duale Form des Zwei-Stichprobenproblems)

Grundidee: Wir betrachten das zweiseitigen Testproblem

$$H_0: \lambda_1 = \lambda_2 \quad \text{gegen} \quad H_1: \lambda_1 \neq \lambda_2$$

$$A = \{(k_1, k_2) \mid \text{Test kann } H_0 \text{ nicht verwerfen}\}$$

= Annahmebereich des Tests zum Signifikanzniveau  $2\alpha$

(mit symmetrischer Aufteilung von  $\alpha$ ).

Linke und rechte Prognosegrenzen:

$$k_\ell^P(k_1, \alpha) = \text{kleinster Wert von } k_2 \text{ mit } k_2 \in A$$

$$k_r^P(k_1, \alpha) = \text{größter Wert von } k_2 \text{ mit } k_2 \in A.$$

Dann gilt:

$$P_\lambda \{k_\ell^P(X_1, \alpha) \leq X_2\} \geq 1 - \alpha \quad \text{für beliebiges } \lambda > 0;$$

$$P_\lambda \{X_2 \leq k_r^P(X_1, \alpha)\} \geq 1 - \alpha \quad \text{für beliebiges } \lambda > 0;$$

$$P_\lambda \{k_\ell^P(X_1, \alpha) \leq X_2 \leq k_r^P(X_1, \alpha)\} \geq 1 - 2\alpha \quad \text{für beliebiges } \lambda > 0.$$

*Beispiel:*

$$k_1 = 100$$

$$1 - 2\alpha = 0.99$$

$$\alpha = 0.005$$

$Y = X_1 \mid X=k \sim \text{Bi}\left(k, \frac{1}{2}\right)$  wobei  $X = X_1 + X_2$

$$k_\ell^P = \underline{k} - k_1 = 166 - 100 = 66$$

$$k_r^P = \bar{k} - k_1 = 240 - 100 = 140$$

Linke und rechte  $p$ -Werte:

$k$	$P\{Y \leq k_1\}$	$P\{Y \geq k_1\}$
$\vdots$		$\vdots$
165		$0.00396 \leq \alpha$
$\underline{k} = 166$		$0.00511 > \alpha$
$\vdots$	$\vdots$	$\vdots$
$\bar{k} = 240$	$0.00583 > \alpha$	
241	$0.00492 \leq \alpha$	
$\vdots$	$\vdots$	

Mit einem Sicherheitsgrad von 99% gilt:

$$66 \leq X_2 \leq 140.$$

### 3. Verallgemeinerung

$$X_1 \sim \text{Po}(c\lambda)$$

$$X_2 \sim \text{Po}(\lambda)$$

$X_1, X_2$  unabhängig

$\lambda$  unbekannt,  $c$  bekannt

$k_1$  Realisation von  $X_1$  (bekannt)

$k_2$  Realisation von  $X_2$  (noch nicht bekannt)

Bestimme aufgrund der Realisation  $k_1$  Prognosegrenzen für  $X_2$  mit einem vorgegebenen Sicherheitsgrad.

Weiter analog zum symmetrischen Fall mit  $c = 1$ .

*Beispiel:*

$$c = 2$$

$$k_1 = 200$$

$$1 - 2\alpha = 0.99$$

$$\alpha = 0.005$$

$$Y = X_1 | X=k \sim \text{Bi}(k, \vartheta_0) \text{ mit } \vartheta_0 = c/(1+c) = \frac{2}{3}$$

wobei  $X = X_1 + X_2$

$$k_\ell^P = \underline{k} - k_1 = 270 - 200 = 70$$

$$k_r^P = \bar{k} - k_1 = 334 - 200 = 134$$

Mit einem Sicherheitsgrad von 99% gilt:

$$70 \leq X_2 \leq 134.$$

Linke und rechte  $p$ -Werte:

$k$	$P\{Y \leq k_1\}$	$P\{Y \geq k_1\}$
$\vdots$		$\vdots$
269		$0.00393 \leq \alpha$
$\underline{k} = 270$		$0.00520 > \alpha$
$\vdots$	$\vdots$	$\vdots$
$\bar{k} = 334$	$0.00552 > \alpha$	
335	$0.00449 \leq \alpha$	
$\vdots$	$\vdots$	

### 4. Näherungsformeln

Für  $c = 1$  gilt:

$$k_\ell^P(k_1, \alpha) \approx k_1 - z_{1-\alpha} \sqrt{2k_1} + \frac{z_{1-\alpha}^2}{2}$$

$$k_r^P(k_1, \alpha) \approx k_1 + z_{1-\alpha} \sqrt{2k_1} + \frac{z_{1-\alpha}^2}{2}$$

Für beliebiges  $c > 0$  gilt:

$$k_\ell^P(k_1, \alpha) \approx \frac{1}{c} \left[ k_1 - z_{1-\alpha} \sqrt{(1+c)k_1} \right]$$

$$k_r^P(k_1, \alpha) \approx \frac{1}{c} \left[ k_1 + z_{1-\alpha} \sqrt{(1+c)k_1} \right]$$

### 3. Binomialverteilung: Zweistichprobenproblem und Prognoseproblem

#### 1. Problemstellung

$$X_1 \sim \text{Bi}(n_1, \vartheta_1)$$

$$X_2 \sim \text{Bi}(n_2, \vartheta_2)$$

$X_1, X_2$  unabhängig

$n_1, n_2$  bekannt,  $\vartheta_1, \vartheta_2$  unbekannt

1a)  $H_0: \vartheta_1 \leq \vartheta_2$  gegen  $H_1: \vartheta_1 > \vartheta_2$

1b)  $H_0: \vartheta_1 \geq \vartheta_2$  gegen  $H_1: \vartheta_1 < \vartheta_2$

1c)  $H_0: \vartheta_1 = \vartheta_2$  gegen  $H_1: \vartheta_1 \neq \vartheta_2$

$k_1, k_2$  Realisationen von  $X_1, X_2$

Entscheide aufgrund der Realisationen  $(k_1, k_2)$ , ob  $H_0$  oder  $H_1$  vorliegt.

#### 2. Lösung zum Testproblem

Grundidee des Tests:

$$P\{X_1 = k_1 | X = k\} = c \binom{n_1}{k_1} \binom{n_2}{k - k_1} \rho^{k_1}$$

$$\text{wobei } \rho = \frac{\vartheta_1(1 - \vartheta_2)}{\vartheta_2(1 - \vartheta_1)} \quad \text{und} \quad c = \sum_{j=0}^k \binom{n_1}{j} \binom{n_2}{k - j} \rho^j$$

Die Verteilung von  $X_1 | X = k$  heißt verallgemeinerte hypergeometrische Verteilung und wird mit  $H(n, n_1, k, \rho)$  bezeichnet. Falls  $\rho = 1$ , so handelt es sich dabei um die gewöhnliche hypergeometrische Verteilung  $H(n, n_1, k)$ . Es gilt:

$$\vartheta_1 \leq \vartheta_2 \Leftrightarrow \rho \leq 1$$

$$\vartheta_1 > \vartheta_2 \Leftrightarrow \rho > 1$$

Testvorschrift zu vorgegebenem Signifikanzniveau  $\alpha$  (exakter Vier-Felder-Test von Fisher):

Daten in Form einer Vier-Felder-Tafel.

Berechne  $k = k_1 + k_2$ .

Bestimme rechte und linke kritische  $\alpha$ -Schranken zu  $H(n, n_1, k)$ :

$$(k_r, \delta_r) \text{ so: } P\{H(n, n_1, k) > k_r\} + \delta_r P\{H(n, n_1, k) = k_r\} = \alpha$$

$$(k_\ell, \delta_\ell) \text{ so: } P\{H(n, n_1, k) < k_\ell\} + \delta_\ell P\{H(n, n_1, k) = k_\ell\} = \alpha.$$

Testfunktion zum Testproblem 1a):

$$\tau(k_1, k_2) = \begin{cases} 1 & \text{falls } k_1 > k_r \\ \delta_r & \text{falls } k_1 = k_r \\ 0 & \text{falls } k_1 < k_r \end{cases}$$

Für  $k_1 = k_2 = k = 0$ :  $\tau(k_1, k_2) \equiv \alpha$ .

Analog für die Probleme 1b) und 1c).

Vier-Felder-Tafel:

	+	-	
$X_1$	$k_1$	$n_1 - k_1$	$n_1$
$X_2$	$k_2$	$n_2 - k_2$	$n_2$
$X$	$k$	$n - k$	$n$

$p$ -Werte (für nichtrandomisierte Testversion):

$$\text{rechter } p\text{-Wert: } p_r = P(H(n, n_1, k) \geq k_1)$$

$$\text{linker } p\text{-Wert: } p_\ell = P(H(n, n_1, k) \leq k_1).$$

### 3. Näherungsverfahren

Vier-Felder-Tafel in nebenstehender Notation:

Berechne die sogenannte Vier-Felder-Testgröße:

$$T = \frac{ad - bc}{\sqrt{efgh}} \sqrt{n}$$

Vier-Felder-Tafel

	+	-	
$X_1$	$a$	$b$	$e$
$X_2$	$c$	$d$	$f$
$X$	$g$	$h$	$n$

Testvorschrift zum Testproblem 1a):

$$\text{Verwirf } H_0, \text{ falls } T > z_{1-\alpha}$$

(approximativer Test zum Signifikanzniveau  $\alpha$ ).

### 4. Zur Optimalität des Tests bei den einseitigen Fragestellungen 1a) und 1b)

1) Der Test ist optimal (UMP- $\alpha$ -Test, uniformly most powerful) auf jeder Schicht

$$S_k = \{(k_1, k_2) \mid k_1 + k_2 = k\}.$$

2) Der Test ist ähnlich auf der Grenze (similar on the boundary):

$$\beta(\vartheta_1, \vartheta_2) \equiv \alpha \quad \text{falls } \vartheta_1 = \vartheta_2.$$

3) Der Test ist unverfälscht (unbiased):

$$\beta(\vartheta_1, \vartheta_2) \leq \alpha \quad \text{falls } H_0 \text{ wahr;}$$

$$\beta(\vartheta_1, \vartheta_2) > \alpha \quad \text{falls } H_1 \text{ wahr.}$$

4) Der Test besitzt die größtmögliche Macht unter allen unverfälschten Tests (UMPU- $\alpha$ -Test, uniformly most powerful unbiased).

### 5. Prognoseproblem

$$X_1 \sim \text{Bi}(n_1, \vartheta)$$

$$X_2 \sim \text{Bi}(n_2, \vartheta)$$

$X_1, X_2$  unabhängig

$n_1, n_2$  bekannt,  $\vartheta$  unbekannt

$k_1$  Realisation von  $X_1$  ( $k_1$  bekannt)

$k_2$  Realisation von  $X_2$  ( $k_2$  noch nicht bekannt)

Bestimme aufgrund der Realisation  $k_1$  Prognosegrenzen für  $X_2$  mit einem vorgegebenen Sicherheitsgrad.

Weiter wie bei der Poissonverteilung.

Beispiel:

$$n_1 = 60$$

$$n_2 = 100$$

$$k_1 = 10$$

$$1 - 2\alpha = 0.99$$

$$\alpha = 0.005$$

$$Y = X_1 | X=k \sim H(n, n_1, k) \text{ wobei } X = X_1 + X_2$$

$$k_\ell^P = \underline{k} - k_1 = 14 - 10 = 4$$

$$k_r^P = \bar{k} - k_1 = 46 - 10 = 36$$

Vier-Felder-Tafel:

	+	-	
$X_1$	10	50	60
$X_2$	$k_2$	.	100
$X$	$\underline{k}$	.	160

Linke und rechte  $p$ -Werte

$k$	$P\{Y \leq k_1\}$	$P\{Y \geq k_1\}$
$\vdots$		$\vdots$
13		$0.00320 \leq \alpha$
$\underline{k} = 14$		$0.00779 > \alpha$
$\vdots$	$\vdots$	$\vdots$
$\bar{k} = 46$	$0.00648 > \alpha$	
47	$0.00454 \leq \alpha$	
$\vdots$	$\vdots$	

Mit einem Sicherheitsgrad von 99% gilt:  $4 \leq X_2 \leq 36$ .

## 6. Näherungsformeln für die Prognosegrenzen

Bei gegebenem  $n_1, n_2, k_1$  und  $\alpha$  gilt:

$$k_\ell^P \approx \frac{n_2}{n_1} \left[ k_1 - z_{1-\alpha} \sqrt{n_1 p_1 (1-p_1) n/n_2} \right]$$

$$k_r^P \approx \frac{n_2}{n_1} \left[ k_1 + z_{1-\alpha} \sqrt{n_1 p_1 (1-p_1) n/n_2} \right]$$

wobei  $n = n_1 + n_2$  und  $p_1 = k_1/n_1$ .

## 7. Vertrauensgrenzen für den Parameter $\rho$ (= Odds Ratio)

$$X_1 \sim \text{Bi}(n_1, \vartheta_1)$$

$$X_2 \sim \text{Bi}(n_2, \vartheta_2)$$

$X_1, X_2$  unabhängig

$n_1, n_2$  bekannt,  $\vartheta_1, \vartheta_2$  unbekannt

$$X_1 | X=k \sim H(n, n_1, k, \rho)$$

$$\rho = \frac{\vartheta_1(1-\vartheta_2)}{\vartheta_2(1-\vartheta_1)} \quad (\text{= Odds Ratio, natürlicher Parameter})$$

Linke Vertrauensgrenze für  $\rho$ :

$$\rho_\ell = \rho_\ell(k_1, k_2, n_1, n_2, \alpha): \quad P\{H(n, n_1, k, \rho_\ell) \geq k_1\} = \alpha$$

$$\rho_r = \rho_r(k_1, k_2, n_1, n_2, \alpha): \quad P\{H(n, n_1, k, \rho_r) \leq k_1\} = \alpha.$$

## 8. Verallgemeinerung

Die Verallgemeinerung des Zwei-Stichprobenproblems bei der Binomialverteilung und des zugehörigen Prognoseproblems auf den Fall

$$\text{a) } H_0: \rho \leq \rho_0 \text{ gegen } H_1: \rho > \rho_0 \text{ wobei } \rho = \frac{\vartheta_1(1-\vartheta_2)}{\vartheta_2(1-\vartheta_1)} = \text{Odds Ratio}$$

erfolgt wie bei der Poissonverteilung im Fall  $c \neq 1$ . Ebenso Fall b) und c).

## 4. Binomialverteilung: Asymptotische Gütefunktion zum Vier-Felder-Test

### 1. Testproblem

$$X_1 \sim \text{Bi}(n_1, \vartheta_1)$$

$$X_2 \sim \text{Bi}(n_2, \vartheta_2)$$

$X_1, X_2$  unabhängig

$n_1, n_2$  bekannt,  $\vartheta_1, \vartheta_2$  unbekannt

$H_0: \vartheta_1 \leq \vartheta_2$  gegen  $H_1: \vartheta_1 > \vartheta_2$

$k_1, k_2$  Realisationen von  $X_1, X_2$

Entscheide aufgrund der Realisationen  $(k_1, k_2)$ , ob  $H_0$  oder  $H_1$  vorliegt.

### 2. Lösung zum Testproblem

Falls  $\vartheta_1 = \vartheta_2$ , so gilt:  $X_1 | X=k \sim H(n, n_1, k)$ .

Testvorschrift zu vorgegebenem Signifikanzniveau  $\alpha$  (exakter Vier-Felder-Test von Fisher):

Daten in Form einer Vier-Felder-Tafel.

Berechne  $k = k_1 + k_2$ .

Bestimme die rechte kritische  $\alpha$ -Schranken zu  $H(n, n_1, k)$ :

$$(k_r, \delta_r) \text{ so: } P\{H(n, n_1, k) > k_r\} + \delta_r P\{H(n, n_1, k) = k_r\} = \alpha$$

Verwirf  $H_0$ , falls  $k_1 > k_r$ .

Vier-Felder-Tafel:

	+	-	
$X_1$	$k_1$	$n_1 - k_1$	$n_1$
$X_2$	$k_2$	$n_2 - k_2$	$n_2$
$X$	$k$	$n - k$	$n$

Da wir uns im Folgenden nur für die asymptotische Gütefunktion interessieren, so können wir uns auf den nicht-randomisierten Test beschränken.

### 3. Asymptotische Gütefunktion im Spezialfall $n_1 = n_2$

Die Gütefunktion des Vier-Felder-Tests ist gegeben durch

$$\beta(\vartheta_1, \vartheta_2) = P_{(\vartheta_1, \vartheta_2)}\{\text{Test verwirft } H_0\}.$$

Wir wollen nun die Stichprobenumfänge  $n_1, n_2$  gegen  $\infty$  streben lassen und den Grenzwert der Gütefunktion bestimmen. Es ist sehr plausibel (und es wird sich herausstellen), dass der Vier-Felder-Test konsistent ist, d.h. dass

$$\beta(\vartheta_1, \vartheta_2) \xrightarrow{n_1, n_2 \rightarrow \infty} \begin{cases} 1 & \text{falls } \vartheta_1 > \vartheta_2 \\ \alpha & \text{falls } \vartheta_1 = \vartheta_2 \\ 0 & \text{falls } \vartheta_1 < \vartheta_2 \end{cases}$$

Mit dieser Beziehung ist die Gütefunktion für große Werte von  $n_1, n_2$  schlecht approximierbar, da die Grenzfunktion eine ausgeartete Gütefunktion ist. Um eine nicht-ausgeartete Grenzfunktion zu finden, haben wir für wachsendes  $n_1, n_2$  die "Auflösung" zu verfeinern.

Wir nehmen jetzt an, daß  $n_1 = n_2 = m$ ; dann ist  $n = n_1 + n_2 = 2m$ .

Falls  $\vartheta_1 = \vartheta_2$ , so gilt:  $X_1 | X=k \sim H(n, m, k)$ . Wir setzen

$d = \vartheta_1 - \vartheta_2$  und haben dann  $H_0: d \leq 0$  gegen  $H_1: d > 0$ . Da der

Test konsistent ist, so konvergiert die Gütefunktion für jedes  $d > 0$

gegen 1. Nun wollen wir die Auflösung in  $d$  verfeinern und führen

anstelle von  $\vartheta_1$  und  $\vartheta_2$  die beiden Parameter

$$\bar{\vartheta} = \frac{1}{2}(\vartheta_1 + \vartheta_2) \quad \text{und} \quad \delta = \sqrt{n}(\vartheta_1 - \vartheta_2)$$

ein. Unser Testproblem kann nun geschrieben werden als  $H_0: \delta \leq 0$  gegen  $H_1: \delta > 0$ , und es gilt

$$\vartheta_1 = \bar{\vartheta} + \frac{\delta}{2\sqrt{n}} \quad \text{und} \quad \vartheta_2 = \bar{\vartheta} - \frac{\delta}{2\sqrt{n}}$$

Wir wollen bei festem  $\alpha$ ,  $\bar{\vartheta}$  und  $\delta$  den Grenzwert der Gütefunktion für  $n \rightarrow \infty$  bestimmen. Bei

festem  $\delta$  wird also der Abstand  $\vartheta_1 - \vartheta_2 = \delta/\sqrt{n}$  immer kleiner, d.h. die "Auflösung" bei der

Gütefunktion wird immer feiner. Wir schreiben jetzt die Gütefunktion als Funktion von  $\delta$ :

$$\beta_n(\delta) = P_{\otimes} \{ \text{Test verwirft } H_0 \}, \quad \text{wobei } \otimes = (n, \alpha, \bar{\vartheta}, \delta),$$

und in dieser Form konvergiert die Gütefunktion für  $n \rightarrow \infty$  gegen eine nicht-ausgeartete Grenzfunktion, die wir bestimmen wollen. Es gilt die Darstellung

$$(1) \quad \beta_n(\delta) = P_{\otimes} \{ \text{Test verwirft } H_0 \} = \sum_{k=0}^n P_{\otimes} \{ X = k \} \underbrace{P_{\otimes} \{ \text{Test verwirft } H_0 | X = k \}}_{= \beta_{n|k}(\delta)},$$

wobei  $\beta_{n|k}(\delta)$  die Gütefunktion des bedingten Tests gegeben  $X = k$  bezeichnet. Offensichtlich gilt

$$(2) \quad \frac{X}{n} = \frac{1}{2} \frac{X_1 + X_2}{m} \xrightarrow{m \rightarrow \infty} \frac{1}{2}(\vartheta_1 + \vartheta_2) = \bar{\vartheta}$$

und somit ist für großes  $n$  der Quotient  $k/n$  mit großer Wahrscheinlichkeit in der Nähe von  $\bar{\vartheta}$ .

Nun gilt

$$\beta_{n|k}(\delta) = P_{\otimes} \{ \text{Test verwirft } H_0 | X = k \} = P_{\otimes} \{ X_1 > k_r | X = k \},$$

und wir stehen vor dem Problem, den Grenzwert dieser bedingten Wahrscheinlichkeit zu be-

stimmen. Dabei ist die bedingte Verteilung von  $X_1$  gegeben  $X = k$  eine verallgemeinerte hypergeometrische Verteilung, falls  $\delta \neq 0$ . Eine einfache Idee hilft, das Problem zu lösen. Es gilt

$$\begin{aligned} \beta_{n|k}(\delta) &= P_{\otimes} \{ X_1 > k_r | X = k \} = P_{\otimes} \{ 2X_1 > 2k_r | X_1 + X_2 = k \} = \\ &= P_{\otimes} \{ X_1 - X_2 > 2k_r - k | X_1 + X_2 = k \} \end{aligned}$$

Wenn es uns gelingt zu zeigen, dass die beiden Zufallsgrößen  $X_1 + X_2$  und  $X_1 - X_2$  stochastisch unabhängig sind, so gilt

$$(3) \quad \beta_{n|k}(\delta) = P_{\otimes} \{ X_1 - X_2 > 2k_r - k | X_1 + X_2 = k \} = P_{\otimes} \{ X_1 - X_2 > 2k_r - k \},$$

und damit ist unser Problem wesentlich vereinfacht.

	+	-	
$X_1$	$k_1$	$m - k_1$	$m$
$X_2$	$k_2$	$m - k_2$	$m$
$X$	$k$	$n - k$	$n$

#### 4. Zur Unabhängigkeit von $X_1 + X_2$ und $X_1 - X_2$

a) Es sei

$Z_1, Z_2$  unabhängig je mit  $N(0,1)$ ;

$$U = Z_1 + Z_2; V = Z_1 - Z_2.$$

Dann gilt

$U, V$  unkorreliert und somit stochastisch unabhängig ..

b) Es sei

$X_1 \sim N(\mu_1, \sigma^2); X_2 \sim N(\mu_2, \sigma^2); X_1, X_2$  unabhängig;

$$U = X_1 + X_2; V = X_1 - X_2.$$

Dann gilt

$$E(U)E(V) = (\mu_1 + \mu_2)(\mu_1 - \mu_2) = \mu_1^2 - \mu_2^2$$

$$E(UV) = E(X_1^2 - X_2^2) = (\sigma^2 + \mu_1^2) - (\sigma^2 + \mu_2^2) = \mu_1^2 - \mu_2^2$$

$$\Rightarrow \text{Cov}(U, V) = E(UV) - E(U)E(V) = 0,$$

d.h.  $U, V$  unkorreliert und somit stochastisch unabhängig .

c) Es sei

$X_1 \sim \text{Bi}(m, \vartheta_1); X_2 \sim \text{Bi}(m, \vartheta_2); X_1, X_2$  unabhängig;

$$Y_1 = \frac{X_1 - m\bar{\vartheta}}{\sqrt{m\bar{\vartheta}(1-\bar{\vartheta})}}; Y_2 = \frac{X_2 - m\bar{\vartheta}}{\sqrt{m\bar{\vartheta}(1-\bar{\vartheta})}}.$$

Dann sind  $Y_1$  und  $Y_2$  stochastisch unabhängig, und es gilt

$$E(Y_1) = \frac{m(\vartheta_1 - \bar{\vartheta})}{\sqrt{m\bar{\vartheta}(1-\bar{\vartheta})}} = \frac{\sqrt{m}}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}} \cdot \frac{\delta}{2\sqrt{n}} = \frac{\delta}{2\sqrt{2}\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}} = \mu_1;$$

$$E(Y_2) = \frac{m(\vartheta_2 - \bar{\vartheta})}{\sqrt{m\bar{\vartheta}(1-\bar{\vartheta})}} = -\frac{\sqrt{m}}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}} \cdot \frac{\delta}{2\sqrt{n}} = -\frac{\delta}{2\sqrt{2}\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}} = \mu_2;$$

$$\text{Var}(Y_1) = \frac{m\vartheta_1(1-\vartheta_1)}{m\bar{\vartheta}(1-\bar{\vartheta})} = \frac{\vartheta_1(1-\vartheta_1)}{\bar{\vartheta}(1-\bar{\vartheta})} \xrightarrow{m \rightarrow \infty} 1 \text{ denn } \vartheta_1 \rightarrow \bar{\vartheta} \text{ für } m \rightarrow \infty;$$

$$\text{Var}(Y_2) = \frac{m\vartheta_2(1-\vartheta_2)}{m\bar{\vartheta}(1-\bar{\vartheta})} = \frac{\vartheta_2(1-\vartheta_2)}{\bar{\vartheta}(1-\bar{\vartheta})} \xrightarrow{m \rightarrow \infty} 1 \text{ denn } \vartheta_2 \rightarrow \bar{\vartheta} \text{ für } m \rightarrow \infty;$$

$$(4) Y_1 \xrightarrow{d} N(\mu_1, 1) \text{ für } m \rightarrow \infty;$$

$$Y_2 \xrightarrow{d} N(\mu_2, 1) \text{ für } m \rightarrow \infty.$$

Nun sei

$$U = Y_1 + Y_2; V = Y_1 - Y_2.$$

Dann gilt aufgrund von b)

$U, V$  unkorreliert für  $m \rightarrow \infty$  und somit stochastisch unabhängig für  $m \rightarrow \infty$  .

## 5. Asymptotische Gütefunktion, Fortsetzung

Die Gütefunktion des bedingten Tests kann aufgrund von (3) geschrieben werden als

$$\begin{aligned}\beta_{n|k}(\delta) &= P_{\otimes} \left\{ X_1 - X_2 > 2k_r - k \mid X_1 + X_2 = k \right\} \\ &= P_{\otimes} \left\{ Y_1 - Y_2 > \frac{1}{\sigma} (2k_r - k) \mid Y_1 + Y_2 = \frac{k + 2\mu}{\sigma} \right\},\end{aligned}$$

wobei

$$Y_1 = \frac{X_1 - \mu}{\sigma} \quad \text{und} \quad Y_2 = \frac{X_2 - \mu}{\sigma} \quad \text{mit} \quad \mu = m\bar{\vartheta}, \sigma^2 = m\bar{\vartheta}(1-\bar{\vartheta}).$$

Nun sind die Zufallsgrößen  $Y_1 + Y_2$  und  $Y_1 - Y_2$  stochastisch unabhängig für  $m \rightarrow \infty$ , und daher gilt für großes  $n = 2m$

$$(5) \quad \beta_{n|k}(\delta) \approx P_{\otimes} \left\{ Y_1 - Y_2 > \frac{1}{\sigma} (2k_r - k) \right\}.$$

Aufgrund von (4) gilt für  $m \rightarrow \infty$

$$Y_1 - Y_2 \xrightarrow{d} N(\mu_1 - \mu_2, 2), \quad \text{wobei} \quad \mu_1 - \mu_2 = \frac{\delta}{\sqrt{2}\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}.$$

Nun haben wir noch den Grenzwert von  $(2k_r - k)/\sigma$  zu bestimmen. Es gilt für  $n = 2m \rightarrow \infty$

$$(6) \quad \frac{X_1 + X_2}{n} \xrightarrow{P} \bar{\vartheta},$$

und somit liegt für großes  $m$  der Quotient  $k/n$  mit großer Wahrscheinlichkeit in der Nähe von  $\bar{\vartheta}$ . Wir betrachten nun den Grenzübergang

$$(7) \quad n (= 2m) \rightarrow \infty, \quad k/n \rightarrow \bar{\vartheta} \quad (\delta, \bar{\vartheta}, \alpha \text{ fest}).$$

$k_r$  ist die rechte kritische  $\alpha$ -Schranke der hypergeometrischen Verteilung  $H(n, m, k)$ ; es sei  $Y \sim H(n, m, k)$ . Dann gilt

$$E(Y) = \mu_Y = \frac{1}{2}k \quad \text{und} \quad \text{Var}(Y) = \sigma_Y^2 = \frac{k(n-k)}{4(n-1)} \approx \frac{k(n-k)}{4n}.$$

Beim Grenzübergang (7) gilt  $\sigma_Y^2 \rightarrow \infty$ . Weiter gilt

$$\frac{Y - \mu_Y}{\sigma_Y} \xrightarrow{d} N(0, 1) \quad \text{für} \quad \sigma_Y^2 \rightarrow \infty,$$

und daher

$$\frac{k_r - \mu_Y}{\sigma_Y} \rightarrow z_{1-\alpha}, \quad \text{für} \quad \sigma_Y^2 \rightarrow \infty.$$

Nun ist

$$\frac{1}{\sigma} (2k_r - k) = \frac{2\sigma_Y}{\sigma} \cdot \frac{k_r - \mu_Y}{\sigma_Y} = \frac{2\sigma_Y}{\sqrt{m\bar{\vartheta}(1-\bar{\vartheta})}} \cdot \frac{k_r - \mu_Y}{\sigma_Y},$$

und es gilt

$$\frac{\sigma_Y^2}{m} = \frac{2\sigma_Y^2}{n} = \frac{k(n-k)}{2n(n-1)} \rightarrow \frac{1}{2}\bar{\vartheta}(1-\bar{\vartheta}) \quad \text{für} \quad k/n \rightarrow \bar{\vartheta},$$

und somit erhalten wir beim Grenzübergang (7)

$$\frac{1}{\sigma} (2k_r - k) = \frac{2\sigma_Y}{\sqrt{m\bar{\vartheta}(1-\bar{\vartheta})}} \cdot \frac{k_r - \mu_Y}{\sigma_Y} \rightarrow \sqrt{2} z_{1-\alpha}.$$

Somit erhalten wir aus (5) beim Grenzübergang (7)

$$\beta_{n|k}(\delta) \rightarrow P \left\{ N(\mu_1 - \mu_2, 2) > \sqrt{2} z_{1-\alpha} \right\} = \beta(\delta).$$

Wenn wir mit  $U$  eine Zufallsgröße mit einer Normalverteilung  $N(\mu_1 - \mu_2, 2)$  bezeichnen, dann kann die Grenzfunktion auch geschrieben werden als

$$\begin{aligned}\beta(\delta) &= P\left\{U > \sqrt{2} z_{1-\alpha}\right\} = P\left\{\frac{U - (\mu_1 - \mu_2)}{\sqrt{2}} > z_{1-\alpha} - \frac{\mu_1 - \mu_2}{\sqrt{2}}\right\} \\ &= P\left\{N(0,1) > z_{1-\alpha} - \frac{\delta}{2\sqrt{\vartheta(1-\vartheta)}}\right\}.\end{aligned}$$

*Zusammenfassung:*

Die Gütefunktion des Vier-Felder-Tests kann gemäß (1) dargestellt werden als

$$\beta_n(\delta) = P_{\otimes}\{\text{Test verwirft } H_0\} = \sum_{k=0}^n P_{\otimes}\{X = k\} \underbrace{P_{\otimes}\{\text{Test verwirft } H_0 | X = k\}}_{= \beta_{n|k}(\delta)},$$

wobei  $\otimes = (n, \alpha, \bar{\vartheta}, \delta)$ . Gemäß (2) gilt

$$\frac{X}{n} \xrightarrow{P} \bar{\vartheta} \text{ für } n \rightarrow \infty,$$

und beim Grenzübergang  $n (= 2m) \rightarrow \infty$ ,  $k/n \rightarrow \bar{\vartheta}$  ( $\delta, \bar{\vartheta}, \alpha$  fest) gilt

$$\beta_{n|k}(\delta) \rightarrow \beta(\delta) = P\left\{N(0,1) > z_{1-\alpha} - \frac{\delta}{2\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}\right\}.$$

Die Grenzfunktion  $\beta(\delta)$  ist stetig in  $\bar{\vartheta}$ , und daraus folgt

$$(8) \quad \beta_n(\delta) \rightarrow \beta(\delta) = P\left\{N(0,1) > z_{1-\alpha} - \frac{\delta}{2\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}}\right\} \text{ für } n \rightarrow \infty \text{ (} \alpha, \bar{\vartheta}, \delta \text{ fest)}.$$

*Folgerung*

Aus (8) ergibt sich insbesondere die *Konsistenz* des Vier-Felder-Tests, denn es gilt

$$\delta = \sqrt{n} d \rightarrow \begin{cases} \infty & \text{für } d = \vartheta_1 - \vartheta_2 > 0 \\ 0 & \text{für } d = \vartheta_1 - \vartheta_2 = 0 \\ -\infty & \text{für } d = \vartheta_1 - \vartheta_2 < 0 \end{cases} \quad \text{und} \quad \beta(\delta) \rightarrow \begin{cases} 1 & \text{für } \delta \rightarrow \infty \\ \alpha & \text{für } \delta \rightarrow 0 \\ 0 & \text{für } \delta \rightarrow -\infty \end{cases}$$

## 6. Asymptotische Gütefunktion im allgemeinen Fall

Wir haben oben die asymptotische Gütefunktion im Spezialfall  $n_1 = n_2 (= m)$  hergeleitet. Hier wollen wir den allgemeinen Fall behandeln. Es sei

$$X_1 \sim \text{Bi}(n_1, \vartheta_1)$$

$$X_2 \sim \text{Bi}(n_2, \vartheta_2)$$

$X_1, X_2$  unabhängig

$n_1, n_2$  bekannt,  $\vartheta_1, \vartheta_2$  unbekannt

$$H_0: \vartheta_1 \leq \vartheta_2 \quad \text{gegen} \quad H_1: \vartheta_1 > \vartheta_2$$

$k_1, k_2$  Realisationen von  $X_1, X_2$

$k_r = k_r(n, n_1, k, \alpha)$  rechte kritische  $\alpha$ -Schranken zu  $H(n, n_1, k)$ .

Der Vier-Felder-Test verwirft  $H_0$ , falls  $k_1 > k_r$ .

Wir wollen die asymptotische Gütefunktion bestimmen für  $n \rightarrow \infty$  und  $n_1/n \rightarrow \lambda$  mit  $0 < \lambda < 1$ .

Es gilt dann  $n_2/n \rightarrow 1 - \lambda$  und weiter

$$X_1/n_1 \rightarrow \vartheta_1$$

$$X_2/n_2 \rightarrow \vartheta_2$$

$$X/n = (X_1 + X_2)/n \rightarrow \lambda\vartheta_1 + (1-\lambda)\vartheta_2 = \bar{\vartheta}$$

Ferner ist

$$\vartheta_1 - \bar{\vartheta} = (1-\lambda)(\vartheta_1 - \vartheta_2)$$

$$\vartheta_2 - \bar{\vartheta} = -\lambda(\vartheta_1 - \vartheta_2)$$

Da der Vier-Felder-Test konsistent ist, so haben wir wiederum die Auflösung der Gütefunktion zu verfeinern mit wachsendem  $n$ , damit wir eine nicht-ausgeartete Grenzfunktion erhalten. Wir setzen daher wiederum  $\delta = \sqrt{n}(\vartheta_1 - \vartheta_2)$  und haben dann

$$\vartheta_1 = \bar{\vartheta} + (1-\lambda)(\vartheta_1 - \vartheta_2) = \bar{\vartheta} + (1-\lambda)\frac{\delta}{\sqrt{n}}$$

$$\vartheta_2 = \bar{\vartheta} - \lambda(\vartheta_1 - \vartheta_2) = \bar{\vartheta} - \lambda\frac{\delta}{\sqrt{n}}$$

Wir wollen nun den Grenzwert der Gütefunktion bestimmen beim Grenzübergang

(9)  $n \rightarrow \infty, n_1/n \rightarrow \lambda$  ( $\bar{\vartheta}, \delta, \alpha, \lambda$  fest).

Wir können jetzt die Gütefunktion schreiben als

$$(10) \quad \beta_n(\delta) = P_{\otimes} \{ \text{Test verwirft } H_0 \} = \sum_{k=0}^n P_{\otimes} \{ X = k \} \underbrace{P_{\otimes} \{ \text{Test verwirft } H_0 | X = k \}}_{= \beta_{n|k}(\delta)},$$

wobei  $\otimes = (n, n_1, \alpha, \bar{\vartheta}, \delta)$ . Da  $X/n \xrightarrow{P} \bar{\vartheta}$ , so haben wir den Grenzwert von  $\beta_{n|k}(\delta)$  zu bestimmen beim Grenzübergang

(11)  $n \rightarrow \infty, n_1/n \rightarrow \lambda, k/n \rightarrow \bar{\vartheta}$  ( $\bar{\vartheta}, \delta, \alpha, \lambda$  fest).

Nun gilt

$$(12) \quad \beta_{n|k}(\delta) = P_{\otimes} \{ \text{Test verwirft } H_0 | X = k \} = P_{\otimes} \{ X_1 > k_r | X = k \},$$

und wir möchten wiederum die bedingte Wahrscheinlichkeit  $P_{\otimes} \{ A | B \}$  vereinfachen, indem wir die Ereignisse  $A$  und  $B$  so umformen, dass sie asymptotisch unabhängig werden. Dazu benötigen wir zwei Hilfssätze zur stochastischen Unabhängigkeit.

Vier-Felder-Tafel:

	+	-	
$X_1$	$k_1$	$n_1 - k_1$	$n_1$
$X_2$	$k_2$	$n_2 - k_2$	$n_2$
$X$	$k$	$n - k$	$n$

Zur stochastischen Unabhängigkeit

a) Es sei

$$X_1 \sim N(\mu_1, \sigma_1^2); X_2 \sim N(\mu_2, \sigma_2^2); X_1, X_2 \text{ unabhängig};$$

$$U = X_1 + X_2; V = aX_1 - bX_2, \text{ wobei } a\sigma_1^2 = b\sigma_2^2.$$

Dann gilt

$$E(U)E(V) = (\mu_1 + \mu_2)(a\mu_1 - b\mu_2) = a\mu_1^2 - b\mu_2^2 + (a-b)\mu_1\mu_2$$

$$\begin{aligned} E(UV) &= E(aX_1^2 - bX_2^2 + (a-b)X_1X_2) = a(\sigma_1^2 + \mu_1^2) - b(\sigma_2^2 + \mu_2^2) + (a-b)\mu_1\mu_2 \\ &= a\mu_1^2 - b\mu_2^2 + (a-b)\mu_1\mu_2 \end{aligned}$$

$$\Rightarrow \text{Cov}(U, V) = E(UV) - E(U)E(V) = 0,$$

d.h.  $U, V$  unkorreliert und somit stochastisch unabhängig .

b) Es sei

$$X_1 \sim \text{Bi}(n_1, \vartheta_1); X_2 \sim \text{Bi}(n_2, \vartheta_2); X_1, X_2 \text{ unabhängig};$$

$$Y_1 = \frac{X_1 - n_1\bar{\vartheta}}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}; Y_2 = \frac{X_2 - n_2\bar{\vartheta}}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}.$$

Dann sind  $Y_1$  und  $Y_2$  stochastisch unabhängig, und es gilt beim Grenzübergang (11)

$$E(Y_1) = \frac{n_1(\vartheta_1 - \bar{\vartheta})}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}} = \frac{n_1}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}} \cdot \frac{(1-\lambda)\delta}{\sqrt{n}} = \frac{n_1}{n} \cdot \frac{(1-\lambda)\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}} = \mu_1 \rightarrow \frac{\lambda(1-\lambda)\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}};$$

$$E(Y_2) = \frac{n_2(\vartheta_2 - \bar{\vartheta})}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}} = -\frac{n_2}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}} \cdot \frac{\lambda\delta}{\sqrt{n}} = -\frac{n_2}{n} \cdot \frac{\lambda\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}} = \mu_2 \rightarrow -\frac{\lambda(1-\lambda)\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}};$$

$$\text{Var}(Y_1) = \frac{n_1\vartheta_1(1-\vartheta_1)}{n\bar{\vartheta}(1-\bar{\vartheta})} \rightarrow \lambda \quad \text{denn } \vartheta_1 \rightarrow \bar{\vartheta};$$

$$\text{Var}(Y_2) = \frac{n_2\vartheta_2(1-\vartheta_2)}{n\bar{\vartheta}(1-\bar{\vartheta})} \Rightarrow 1-\lambda \quad \text{denn } \vartheta_2 \rightarrow \bar{\vartheta};$$

$$(13) \quad Y_1 \xrightarrow{d} N(\mu_1, \lambda);$$

$$Y_2 \xrightarrow{d} N(\mu_2, 1-\lambda).$$

Nun sei

$$U = Y_1 + Y_2; V = (1-\lambda)Y_1 - \lambda Y_2.$$

Dann gilt aufgrund von a)

$U, V$  asymptotisch unkorreliert und somit asymptotisch unabhängig beim Grenzübergang (11).

Jetzt wollen wir die Ergebnisse unter b) anwenden bei unserer Gütefunktion (12)

$$\beta_{n|k}(\delta) = P_{\otimes} \{ \text{Test verwirft } H_0 | X = k \} = P_{\otimes} \{ X_1 > k_r | X_1 + X_2 = k \}.$$

Die Bedingung  $X_1 + X_2 = k$  kann auch geschrieben werden als

$$(14) \quad Y_1 + Y_2 = \frac{k - n\bar{\vartheta}}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}$$

und die Ungleichung  $X_1 > k_r$  als

$$(15) \quad Y_1 > \frac{k_r - n_1\bar{\vartheta}}{\sqrt{n\bar{\vartheta}(1-\bar{\vartheta})}}.$$

Wenn wir von dieser Ungleichung  $\lambda(Y_1 + Y_2)$  subtrahieren und die Gleichung (14) berücksichtigen, so erhalten wir die Ungleichung

$$(1-\lambda)Y_1 - \lambda Y_2 > \frac{k_r - n_1 \bar{\vartheta}}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}} - \lambda \frac{k - n \bar{\vartheta}}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}},$$

und weil die beiden Zufallsgrößen

$$U = (1-\lambda)Y_1 - \lambda Y_2 \quad \text{und} \quad V = Y_1 + Y_2$$

asymptotisch unabhängig sind, so gilt für großes  $n$  beim Grenzprozess (11)

$$(16) \beta_{n|k}(\delta) = P_{\otimes} \{X_1 > k_r | X = k\} \approx P_{\otimes} \{(1-\lambda)Y_1 - \lambda Y_2 > A\},$$

wobei

$$A = \frac{k_r - n_1 \bar{\vartheta}}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}} - \lambda \frac{k - n \bar{\vartheta}}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}}.$$

Nun gilt aufgrund von (13)

$$(1-\lambda)Y_1 - \lambda Y_2 \xrightarrow{d} N(\mu, \sigma^2),$$

wobei

$$\mu = (1-\lambda)\mu_1 - \lambda\mu_2 \rightarrow \frac{\lambda(1-\lambda)\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}};$$

$$\sigma^2 = (1-\lambda)^2 \underbrace{\text{Var}(Y_1)}_{\rightarrow \lambda} + \lambda^2 \underbrace{\text{Var}(Y_2)}_{\rightarrow 1-\lambda} \rightarrow \lambda(1-\lambda).$$

Nun haben wir noch den Grenzwert der Größe  $A$  zu bestimmen.  $k_r$  ist die rechte kritische  $\alpha$ -Schranke zur hypergeometrischen Verteilung  $H(n, n_1, k)$ , und wenn  $Y$  eine Zufallsgröße mit dieser Verteilung bezeichnet, dann gilt

$$E(Y) = \mu_Y = k \frac{n_1}{n} \quad \text{und} \quad \text{Var}(Y) = \sigma_Y^2 = k \frac{n_1}{n} \cdot \frac{n_2}{n} \cdot \frac{n-k}{n-1},$$

und weiter

$$\frac{k_r - \mu_Y}{\sigma_Y} \rightarrow z_{1-\alpha} \quad \text{für} \quad \sigma_Y^2 \rightarrow \infty;$$

beim Grenzübergang (11) gilt offensichtlich  $\sigma_Y^2 \rightarrow \infty$ . Wir können den Ausdruck  $A$  schreiben als

$$(17) A = \underbrace{\frac{k_r - \mu_Y}{\sigma_Y}}_{\rightarrow z_{1-\alpha}} \cdot \frac{\sigma_Y}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}} + \frac{\mu_Y - \lambda k}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}} - \frac{(n_1 - n \lambda) \bar{\vartheta}}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}} = A_1 + A_2 - A_3.$$

Beim Grenzübergang (11) gilt

$$\sigma_Y^2/n \rightarrow \lambda(1-\lambda)\bar{\vartheta}(1-\bar{\vartheta})$$

und somit

$$\frac{\sigma_Y}{\sqrt{n \bar{\vartheta} (1-\bar{\vartheta})}} \rightarrow \sqrt{\lambda(1-\lambda)}.$$

Wir haben also  $A_1 \rightarrow z_{1-\alpha} \sqrt{\lambda(1-\lambda)}$ . Weiter gilt

$$(18) \frac{\mu_Y - \lambda k}{\sqrt{n}} = \frac{k}{\sqrt{n}} \left( \frac{n_1}{n} - \lambda \right).$$

Es gilt zwar  $n_1/n \rightarrow \lambda$ , doch da  $k/\sqrt{n} \rightarrow \infty$  divergiert, können wir den Grenzwert von (18) nicht bestimmen ohne weitere Annahmen über die Konvergenzgeschwindigkeit. Wir wollen daher zusätzlich zur Annahme  $n_1/n \rightarrow \lambda$  noch voraussetzen, daß

$$(19) \sqrt{n} \left( \frac{n_1}{n} - \lambda \right) \rightarrow 0.$$

Dies ist z.B. erfüllt, wenn  $n_1$  bei gegebenem  $\lambda$  und  $n$  als die nächste ganze Zahl bei  $\lambda n$  gewählt wird. Dann ist  $n_1 - \lambda n \leq \frac{1}{2}$  und  $\sqrt{n}(n_1/n - \lambda) \leq 1/(2\sqrt{n}) \rightarrow 0$ . Unter der zusätzlichen Annahme (19) gilt dann beim Grenzübergang (11)

$$\frac{\mu_Y - \lambda k}{\sqrt{n}} = \underbrace{\frac{k}{n}}_{\rightarrow \bar{\vartheta}} \cdot \underbrace{\sqrt{n}\left(\frac{n_1}{n} - \lambda\right)}_{\rightarrow 0} \rightarrow 0,$$

und damit konvergiert  $A_2$  gegen 0. Ebenso konvergiert  $A_3$  gegen 0, denn

$$\frac{(n_1 - n\lambda)}{\sqrt{n}} = \sqrt{n}\left(\frac{n_1}{n} - \lambda\right) \rightarrow 0.$$

Somit gilt  $A \rightarrow z_{1-\alpha}\sqrt{\lambda(1-\lambda)}$ , und wir erhalten aus (16) beim Grenzübergang (11) unter der zusätzlichen Annahme (19)

$$\beta_{n|k}(\delta) = P_{\otimes}\{X_1 > k_r | X = k\} \rightarrow P\{N(\mu, \sigma^2) > z_{1-\alpha}\sqrt{\lambda(1-\lambda)}\} = \beta(\delta),$$

wobei

$$\mu = \frac{\lambda(1-\lambda)\delta}{\sqrt{\bar{\vartheta}(1-\bar{\vartheta})}} \quad \text{und} \quad \sigma^2 = \lambda(1-\lambda).$$

Indem wir die Zufallsgröße  $N(\mu, \sigma^2)$  standardisieren, erhalten wir

$$\beta(\delta) = P\left\{N(0,1) > z_{1-\alpha} - \delta\sqrt{\frac{\lambda(1-\lambda)}{\bar{\vartheta}(1-\bar{\vartheta})}}\right\}.$$

*Zusammenfassung:*

Die Gütefunktion des Vier-Felder-Tests kann gemäß (10) dargestellt werden als

$$\beta_n(\delta) = P_{\otimes}\{\text{Test verwirft } H_0\} = \sum_{k=0}^n P_{\otimes}\{X = k\} \underbrace{P_{\otimes}\{\text{Test verwirft } H_0 | X = k\}}_{= \beta_{n|k}(\delta)},$$

wobei  $\otimes = (n, n_1, \alpha, \bar{\vartheta}, \delta)$ . Es gilt

$$\frac{X}{n} \xrightarrow{P} \bar{\vartheta} \quad \text{für } n \rightarrow \infty,$$

und beim Grenzübergang

$$n \rightarrow \infty, \quad n_1/n \rightarrow \lambda, \quad \sqrt{n}(n_1/n - \lambda) \rightarrow 0, \quad k/n \rightarrow \bar{\vartheta} \quad (\bar{\vartheta}, \delta, \alpha, \lambda \text{ fest})$$

gilt

$$\beta_{n|k}(\delta) \rightarrow \beta(\delta) = P\left\{N(0,1) > z_{1-\alpha} - \delta\sqrt{\frac{\lambda(1-\lambda)}{\bar{\vartheta}(1-\bar{\vartheta})}}\right\}.$$

Die Grenzfunktion  $\beta(\delta)$  ist stetig in  $\bar{\vartheta}$ , und daraus folgt

$$(20) \quad \beta_n(\delta) \rightarrow \beta(\delta) = P\left\{N(0,1) > z_{1-\alpha} - \delta\sqrt{\frac{\lambda(1-\lambda)}{\bar{\vartheta}(1-\bar{\vartheta})}}\right\}$$

für  $n \rightarrow \infty$ ,  $n_1/n \rightarrow \lambda$ ,  $\sqrt{n}(n_1/n - \lambda) \rightarrow 0$  ( $\bar{\vartheta}, \delta, \alpha, \lambda$  fest). Dabei ist

$$\bar{\vartheta} = \lambda\vartheta_1 + (1-\lambda)\vartheta_2 \quad \text{und} \quad \delta = \sqrt{n}(\vartheta_1 - \vartheta_2).$$

Als Folgerung aus (20) ergibt sich die Konsistenz des Vier-Felder-Tests (wie im Spezialfall  $n_1 = n_2$ ).

## 7. Notwendiger Stichprobenumfang

Es sei

$$X_1 \sim \text{Bi}(n_1, \vartheta_1)$$

$$X_2 \sim \text{Bi}(n_2, \vartheta_2)$$

$X_1, X_2$  unabhängig

$$H_0: \vartheta_1 \leq \vartheta_2 \quad \text{gegen} \quad H_1: \vartheta_1 > \vartheta_2$$

$k_1, k_2$  Realisationen von  $X_1, X_2$

$k_r = k_r(n, n_1, k, \alpha)$  rechte kritische  $\alpha$ -Schranken zu  $H(n, n_1, k)$ .

Der Vier-Felder-Test verwirft  $H_0$ , falls  $k_1 > k_r$ .

Vier-Felder-Tafel:

	+	-	
$X_1$	$k_1$	$n_1 - k_1$	$n_1$
$X_2$	$k_2$	$n_2 - k_2$	$n_2$
$X$	$k$	$n - k$	$n$

Nun wollen wir mit Hilfe der asymptotischen Gütefunktion die Stichprobenumfänge  $n_1, n_2$  so bestimmen, daß die Gütefunktion bei einer gegebenen Alternative eine vorgegebene Ablehnwahrscheinlichkeit erreicht.

Gegeben:

$$\vartheta_1, \vartheta_2 \quad \text{mit} \quad \vartheta_1 > \vartheta_2;$$

$$\lambda \quad (= n_1/n).$$

Wir setzen

$$d = \vartheta_1 - \vartheta_2 \quad \text{und} \quad \bar{\vartheta} = \lambda\vartheta_1 + (1-\lambda)\vartheta_2 \approx \frac{1}{2}(\vartheta_1 + \vartheta_2).$$

Nun soll bei gegebenem  $\lambda$  und  $\bar{\vartheta}$  der Stichprobenumfang  $n = n_1 + n_2$  so gewählt werden, daß die Ablehnwahrscheinlichkeit  $> 1 - \alpha_1$  wird bei der Alternative  $(\vartheta_1, \vartheta_2)$ . Wir setzen  $\delta = d\sqrt{n}$  und haben dann aufgrund von (20) die Approximation

$$\begin{aligned} \beta_n(\delta) &\approx \beta(\delta) = P\left\{N(0,1) > z_{1-\alpha} - \delta \sqrt{\frac{\lambda(1-\lambda)}{\bar{\vartheta}(1-\bar{\vartheta})}}\right\} \\ &= P\left\{N(0,1) > z_{1-\alpha} - d\sqrt{n} \sqrt{\frac{\lambda(1-\lambda)}{\bar{\vartheta}(1-\bar{\vartheta})}}\right\} \stackrel{!}{\geq} 1 - \alpha_1 \end{aligned}$$

und daraus ergibt sich

$$z_{1-\alpha} - d\sqrt{n} \sqrt{\frac{\lambda(1-\lambda)}{\bar{\vartheta}(1-\bar{\vartheta})}} \leq -z_{1-\alpha_1}$$

und weiter

$$(21) \quad n \geq \left(\frac{z_{1-\alpha} + z_{1-\alpha_1}}{d}\right)^2 \frac{\bar{\vartheta}(1-\bar{\vartheta})}{\lambda(1-\lambda)}; \quad n_1 = \lambda n; \quad n_2 = (1-\lambda)n.$$

Falls die Werte von  $(\vartheta_1, \vartheta_2)$  unbekannt sind, so kann man in der Formel (21) den Ausdruck  $\bar{\vartheta}(1-\bar{\vartheta})$  durch seine obere Grenze  $1/4$  ersetzen.

*Beispiel*

Es sei  $(\vartheta_1, \vartheta_2) = (0.6, 0.4)$ ,  $\lambda = \frac{1}{3}$ ,  $\alpha = \alpha_1 = 0.025$ . Dann ist  $z_{1-\alpha} = z_{1-\alpha_1} \approx 2$ ,  $d = 0.2$ ,  $\bar{\vartheta} \approx \frac{1}{2}$ , und wir erhalten

$$n \geq \left( \frac{2+2}{0.2} \right)^2 \frac{1/4}{2/9} = 450,$$

und wir wählen daher  $n = 450$ ,  $n_1 = \lambda n = 150$ ,  $n_2 = (1 - \lambda)n = 300$ .

*Kontrolle*

Wir wollen überprüfen, ob das Ergebnis in unserem Beispiel plausibel ist. Beim Parameterpaar  $(\vartheta_1, \vartheta_2) = (0.6, 0.4)$  gilt

$$E(X) = E(X_1) + E(X_2) = n_1\vartheta_1 + n_2\vartheta_2 = 90 + 120 = 210,$$

und die Vier-Felder-Tafel besitzt dann die nebenstehenden Randsummen. Die rechte kritische  $\alpha$ -Schranke ( $\alpha = 0.025$ ) bei der

	+	-	
$X_1$	*	*	150
$X_2$	*	*	300
$X$	210	240	450

hypergeometrischen Verteilung  $H(450, 150, 210)$  beträgt  $k_r = 80$ . Bei der gegebenen Alternative hat der Parameter  $\rho$  (= Odds Ratio) den Wert

$$\rho = \frac{\vartheta_1(1-\vartheta_2)}{(1-\vartheta_1)\vartheta_2} = \frac{0.6 \times 0.6}{0.4 \times 0.4} = 2.25,$$

und die Ablehnwahrscheinlichkeit des bedingten Tests gegeben  $X = k = 210$  beträgt

$$P\{H(n, n_1, k, \rho) > k_r\} = P\{H(450, 150, 210, 2.25) > 80\} = 0.9738 \ (\approx 1 - \alpha_1 = 0.975).$$

Diese Wahrscheinlichkeit kann man berechnen z.B. mit SAS (vgl. HYPERGEO.SAS). Für  $\rho = 0$  beträgt die Ablehnwahrscheinlichkeit  $P\{H(n, n_1, k) > k_r\} = 0.0177$  ( $< \alpha = 0.025$ ).

## 5. Normalverteilung: Vergleich von zwei Mittelwerten

### 1. Problemstellung

$$\left. \begin{array}{l} X_1, \dots, X_{n_1} \\ Y_1, \dots, Y_{n_2} \end{array} \right\} \text{ unabhängige Zufallsgrößen } \begin{cases} \text{je mit } N(\mu_1, \sigma_1^2) \\ \text{je mit } N(\mu_2, \sigma_2^2) \end{cases}$$

$\mu_1, \mu_2$  unbekannt.

1a)  $H_0: \mu_1 \leq \mu_2$  gegen  $H_1: \mu_1 > \mu_2$

1b)  $H_0: \mu_1 \geq \mu_2$  gegen  $H_1: \mu_1 < \mu_2$

1c)  $H_0: \mu_1 = \mu_2$  gegen  $H_1: \mu_1 \neq \mu_2$

### 2. Lösung

#### a) Varianzen bekannt

Falls  $\mu_1 = \mu_2$  und  $\sigma_1, \sigma_2$  bekannt, so gilt

$$D = \bar{X} - \bar{Y} \sim N(0, \sigma_D^2) \quad \text{wobei} \quad \sigma_D^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\frac{D}{\sigma_D} = \frac{\bar{X} - \bar{Y}}{\sigma_D} \sim N(0, 1).$$

Testvorschrift zum Testproblem 1a):

$$\text{Berechne } T = \frac{\bar{X} - \bar{Y}}{\sigma_D} \quad (\sigma_D \text{ wie oben}).$$

Verwirf  $H_0$ , falls  $T > z_{1-\alpha}$ .

*Bemerkung.*

Das Testproblem 1a) kann verallgemeinert werden zu

$$H_0: \mu_1 - \mu_2 \leq \Delta_0 \quad \text{gegen} \quad H_1: \mu_1 - \mu_2 > \Delta_0 \quad \text{wobei } \Delta_0 \text{ bekannt ist.}$$

(Analog bei 1b) und 1c)). Falls  $\mu_1 - \mu_2 = \Delta_0$ , so gilt

$$\frac{\bar{X} - \bar{Y} - \Delta_0}{\sigma_D} \sim N(0, 1) \quad (\sigma_D \text{ wie oben}).$$

Weiter wie oben.

Vertrauensgrenzen für  $\Delta = \mu_1 - \mu_2$ :

$$\Delta_\ell = \bar{X} - \bar{Y} - z_{1-\alpha} \sigma_D \quad (\sigma_D \text{ wie oben}),$$

$$\Delta_r = \bar{X} - \bar{Y} + z_{1-\alpha} \sigma_D.$$

### b) Varianzen unbekannt, jedoch beide gleich

Hier nehmen wir an, daß  $\sigma_1 = \sigma_2 = \sigma$  mit unbekanntem  $\sigma$  ( $\sigma$  ist ein sogenannter Störparameter, englisch: nuisance parameter). Falls  $\mu_1 = \mu_2$ , so gilt

$$T = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sim t(n_1 + n_2 - 2)$$

wobei

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left[ \sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 \right].$$

Testvorschrift zum Testproblem 1a):

Verwirf  $H_0$ , falls  $T > t_{1-\alpha}(n_1 + n_2 - 2)$

(doppelter t-Test).

*Bemerkung.*

Das Testproblem 1a) kann verallgemeinert werden zu

$H_0: \mu_1 - \mu_2 \leq \Delta_0$  gegen  $H_1: \mu_1 - \mu_2 > \Delta_0$  wobei  $\Delta_0$  bekannt ist .

(Analog bei 1b) und 1c)). Falls  $\mu_1 - \mu_2 = \Delta_0$ , so gilt

$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sim t(n_1 + n_2 - 2).$$

Weiter wie oben.

Vertrauensgrenzen für  $\Delta = \mu_1 - \mu_2$ :

$$\Delta_l = \bar{X} - \bar{Y} - t_{1-\alpha}(n_1 + n_2 - 2) S \sqrt{\frac{n_1 + n_2}{n_1 n_2}};$$

$$\Delta_r = \bar{X} - \bar{Y} + t_{1-\alpha}(n_1 + n_2 - 2) S \sqrt{\frac{n_1 + n_2}{n_1 n_2}}.$$

### c) Varianzen unbekannt (Behrens-Fisher-Problem)

Jetzt nehmen wir an, daß über die beiden Varianzen  $\sigma_1^2$  und  $\sigma_2^2$  nichts bekannt ist (zwei Störparameter). Das Testproblem heißt dann Behrens-Fisher-Problem. Es existiert keine klassische exakte Lösung, welche auf der suffizienten Statistik  $(\bar{X}, \bar{Y}, S_1^2, S_2^2)$  beruht.

Lösungsvorschläge:

- 1) Näherungslösung von Welch (1947);
- 2) Lösung mit Vertrauensgrenzen und Bonferroni-Ungleichung.

*Zur Bonferroni-Ungleichung:*

Es seien A, B beliebige zufällige Ereignisse mit

$$P(A) \geq 1 - \alpha,$$

$$P(B) \geq 1 - \beta.$$

Dann gilt:

$$P(A \cap B) \geq 1 - \alpha - \beta \quad (\text{Bonferroni-Ungleichung}).$$

Zur Lösung des Behrens-Fisher-Problems mit Vertrauensgrenzen:

Problemstellung:

$$H_0: \mu_1 \leq \mu_2 \quad \text{gegen} \quad H_1: \mu_1 > \mu_2.$$

Berechne

$$\bar{X}, S_1^2, \mu_1^l = \text{linke Vertrauensgrenze für } \mu_1 \text{ zum Sicherheitsgrad } 1 - \alpha;$$

$$\bar{Y}, S_2^2, \mu_2^r = \text{rechte Vertrauensgrenze für } \mu_2 \text{ zum Sicherheitsgrad } 1 - \alpha.$$

Verwirf  $H_0$ , falls  $\mu_1^l \geq \mu_2^r$ .

Dann ist die Irrtumswahrscheinlichkeit 1. Art  $\leq 2\alpha$  (korrekte Ungleichung).

### 3. Optimale Aufteilung des Stichprobenumfangs bei gegebenen Varianzen

Problemstellung wie bei 1 mit bekannten Varianzen  $\sigma_1^2$  und  $\sigma_2^2$  und mit vorgegebenem Gesamtstichprobenumfang  $n = n_1 + n_2$ .

Frage: Wie sind die Stichprobenumfänge  $n_1$  und  $n_2$  zu wählen, damit der Test möglichst trennscharf wird?

Antwort:  $\frac{n_1}{n_2} = \frac{\sigma_1}{\sigma_2}$

Effizienzvergleich:

Verfahren A: Symmetrische Aufteilung, d.h.  $n_1 = n_2$  (Gesamtumfang  $n_A$ ).

Verfahren B: Optimale Aufteilung, d.h.  $n_1 : n_2 = \sigma_1 : \sigma_2$  (Gesamtumfang  $n_B$ ).

Relative Effizienz von Verfahren A verglichen mit Verfahren B:

$$\text{re}(A:B) = \frac{n_B}{n_A} = \frac{(\sigma_1 + \sigma_2)^2}{2(\sigma_1^2 + \sigma_2^2)}.$$

Es gilt:  $\frac{1}{2} < \text{re}(A:B) \leq 1$ .

Beispiel:

Verfahren A mit  $n_A = 100$  (d.h. mit  $n_1 = n_2 = 50$ ) besitzt die gleiche Gütefunktion wie Verfahren B mit den in der nachstehenden Tabelle angegebenen Umfängen  $n_1$  und  $n_2$

$\sigma_1 : \sigma_2$	$\text{re}(A:B)$	$n_B$	$n_1$	$n_2$
1 : 1	1	100	50	50
2 : 1	0.9	90	60	30
3 : 1	0.8	80	60	20
7 : 1	0.64	64	56	8
51 : 1	$\approx 0.52$	52	51	1

## 6. Normalverteilung: Vergleich von k Mittelwerten; einfache Varianzanalyse (ANOVA)

### 1. Problemstellung

$$\left. \begin{array}{l} Y_{11}, \dots, Y_{1n_1} \\ \dots \\ Y_{k1}, \dots, Y_{kn_k} \end{array} \right\} \text{unabhängige Zufallsgrößen} \left\{ \begin{array}{l} \text{je mit } N(\mu_1, \sigma^2) \\ \dots \\ \text{je mit } N(\mu_k, \sigma^2) \end{array} \right.$$

$\mu_1, \dots, \mu_k, \sigma^2$  unbekannt

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$  gegen  $H_1: \text{nicht alle } \mu_i \text{ gleich.}$

Andere Darstellung des Problems:

$$Y_{ij} = \mu + \alpha_i + E_{ij};$$

alle  $E_{ij}$  stochastisch unabhängig,  $E_{ij} \sim N(0, \sigma^2)$ ;

$\mu, \alpha_1, \dots, \alpha_k, \sigma^2$  unbekannt,  $\alpha_1 + \dots + \alpha_k = 0$ ;

$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$  gegen  $H_1: \text{nicht alle } \alpha_i = 0.$

Zusammenhang mit den obigen Parametern:

$$\mu = \frac{1}{n}(\mu_1 + \dots + \mu_k); \quad \alpha_i = \mu_i - \mu = \text{Effekt in Gruppe } i.$$

Typische Anwendung: Ein *Faktor* mit  $k$  verschiedenen Ausprägungen (z.B. ein Medikament in  $k$  verschiedenen Dosierungen);  $\alpha_1, \dots, \alpha_k$  *Faktoreffekte*

### 2. Lösung

$$\bar{Y} = \frac{1}{n} \sum_{i,j} Y_{ij} \quad \text{Gesamtmittel}$$

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{arithmetisches Mittel in Gruppe } i$$

$$SST = \sum_{i,j} (Y_{ij} - \bar{Y})^2 = \text{sum of squares, total (Summe der quadratischen Abweichungen vom Gesamtmittel)}$$

$$SSF = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2 = \text{sum of squares, factor (Summe der quadratischen Abweichungen der Gruppenmittel vom Gesamtmittel)}$$

$$SSE = \sum_{i,j} (Y_{ij} - \bar{Y}_i)^2 = \text{sum of squares, error (Summe der quadratischen Abweichungen der Beobachtungswerte von den Gruppenmitteln)}$$

Es gilt die sogenannte "Varianzzerlegung" (Analysis of variance, ANOVA) (vgl. Deskriptive Statistik, Varianzzerlegung bei Gruppenbildung):

$$SST = SSF + SSE.$$

Weiter gilt

$$\frac{SSE}{\sigma^2} \sim \chi^2(n-k),$$

und unter  $H_0$  gilt ebenfalls

$$\frac{SST}{\sigma^2} \sim \chi^2(n-1) \quad \text{und} \quad \frac{SSF}{\sigma^2} \sim \chi^2(k-1).$$

Daher ist die Größe

$$T = \frac{SSF / \left[ (k-1)\sigma^2 \right]}{SSE / \left[ (n-k)\sigma^2 \right]} = \frac{SSF / (k-1)}{SSE / (n-k)} \sim F(k-1, n-k).$$

Unter  $H_1$  tendiert die Komponente  $SSF$  zu größeren Werten als unter  $H_0$ , während die Verteilung der Komponente  $SSE$  unter  $H_0$  und  $H_1$  gleich bleibt. Daher lehnt man die Nullhypothese ab, falls

$$\frac{SSF / (k-1)}{SSE / (n-k)} \geq F_{1-\alpha}(k-1, n-k) = \text{oberes } \alpha\text{-Quantil der F-Verteilung mit } (k-1, n-k) \text{ Freiheitsgraden}$$

Die wichtigsten Größen der einfachen Varianzanalyse (= Zerlegung der Quadratsumme) werden üblicherweise wie folgt zusammengestellt:

Source	SS	DF	MS	p-value
Factor	$SSF$	$k-1$	$SSF / (k-1)$	–
Error	$SSE$	$n-k$	$SSE / (n-k)$	$p_{crit}$
Total	$SST$	$n-1$	–	–

Dabei ist

<i>Source</i>	Art der Quadratsumme (Quelle, woher stammt die Komponente)
<i>SS</i>	Komponente der Quadratsumme (SS = Sum of Squares)
<i>DF</i>	Anzahl der Freiheitsgrade (DF = degrees of freedom)
<i>MS</i>	durchschnittliche Quadratsumme (MS = mean squares)

Der p-Wert  $p_{crit}$  wird berechnet nach der Formel

$$p_{crit} = P\{F(k-1, n-k) > t_{obs}\},$$

wobei

$$t_{obs} = \frac{SSF / (k-1)}{SSE / (n-k)}.$$

### 3. Spezialfall $k = 2$ (Zwei-Stichprobenproblem)

$$\left. \begin{array}{l} Y_{11}, \dots, Y_{1n_1} \\ Y_{21}, \dots, Y_{2n_2} \end{array} \right\} \text{ unabhängige Zufallsgrößen } \left\{ \begin{array}{l} \text{je mit } N(\mu_1, \sigma^2) \\ \text{je mit } N(\mu_2, \sigma^2) \end{array} \right.$$

$\mu_1, \mu_2, \sigma^2$  unbekannt

$H_0: \mu_1 = \mu_2$  gegen  $H_1: \mu_1 \neq \mu_2$ .

Nun haben wir zwei Lösungen für dieses Testproblem:

- A) den doppelten t-Test
- B) die einfache Varianzanalyse mit  $k = 2$  (ANOVA).

Wir wollen zeigen daß die einfache Varianzanalyse im Spezialfall  $k = 2$  äquivalent ist zum doppelten t-Test mit der zweiseitigen Fragestellung. Es gilt

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{arithmetisches Mittel in Gruppe } i \quad (i = 1, 2)$$

$$\bar{Y} = \frac{1}{n} \sum_{i,j} Y_{ij} = \frac{1}{n} (n_1 \bar{Y}_1 + n_2 \bar{Y}_2) \quad \text{Gesamtmittel}$$

$$\bar{Y}_1 - \bar{Y} = \frac{1}{n} [n\bar{Y}_1 - (n_1\bar{Y}_1 + n_2\bar{Y}_2)] = \frac{n_2}{n} (\bar{Y}_1 - \bar{Y}_2)$$

$$\bar{Y}_2 - \bar{Y} = \frac{1}{n} [n\bar{Y}_2 - (n_1\bar{Y}_1 + n_2\bar{Y}_2)] = \frac{n_1}{n} (\bar{Y}_2 - \bar{Y}_1)$$

$$SSF = n_1 (\bar{Y}_1 - \bar{Y})^2 + n_2 (\bar{Y}_2 - \bar{Y})^2 = \frac{1}{n^2} (n_1 n_2^2 + n_2 n_1^2) (\bar{Y}_1 - \bar{Y}_2)^2 = \frac{n_1 n_2}{n} (\bar{Y}_1 - \bar{Y}_2)^2$$

$$SSE = \sum_{i,j} (Y_{ij} - \bar{Y}_i)^2 = \sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2 + \sum_{j=2}^{n_2} (Y_{2j} - \bar{Y}_2)^2$$

$$T = \frac{SSF/(k-1)}{SSE/(n-k)} = \frac{SSF}{SSE/(n-2)} = \frac{n_1 n_2}{n} \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{S^2}, \text{ wobei } S^2 = \frac{SSE}{n-2},$$

und dies ist das Quadrat der Testgröße beim t-Test. Wenn  $X$  eine t-Verteilung  $t(n-2)$  besitzt, dann besitzt  $X^2$  eine F-Verteilung  $F(1, n-2) = F(k-1, n-k)$ , und damit ist der F-Test bei der einfachen Varianzanalyse äquivalent zum zweiseitigen doppelten t-Test.

*Bemerkung:*

Beim Mehrstichprobenproblem der einfachen Varianzanalyse gibt es nur die zweiseitige Fragestellung  $H_0: \mu_1 = \mu_2 = \dots = \mu_k$  gegen  $H_1$ : nicht alle  $\mu_i$  gleich. Diese Fragestellung ist nicht sehr informativ; die Nullhypothese kann statistisch nicht bewiesen werden, und die Alternative ist eine "Allerweltsalternative", welche sogleich nach weiteren Fragen ruft: wenn nicht alle  $\mu_i$  gleich sind, so möchte man wissen, in welchen Gruppen die Effekte  $\alpha_i = \mu_i - \mu \neq 0$  sind. Die Alternative muss dann verfeinert werden zu einer mehrfachen Alternative, und dies führt zu *multiplen* Testproblemen.

## 7. Normalverteilung: Vergleich von zwei Varianzen

### 1. Problemstellung

$$\left. \begin{array}{l} X_1, \dots, X_{n_1} \\ Y_1, \dots, Y_{n_2} \end{array} \right\} \text{ unabhängige Zufallsgrößen } \begin{cases} \text{je mit } N(\mu_1, \sigma_1^2) \\ \text{je mit } N(\mu_2, \sigma_2^2) \end{cases}$$

$\sigma_1, \sigma_2$  unbekannt.

1a)  $H_0: \sigma_1 \leq \sigma_2$  gegen  $H_1: \sigma_1 > \sigma_2$ ;

1b)  $H_0: \sigma_1 \geq \sigma_2$  gegen  $H_1: \sigma_1 < \sigma_2$ ;

1c)  $H_0: \sigma_1 = \sigma_2$  gegen  $H_1: \sigma_1 \neq \sigma_2$ .

### 2. Lösung

#### a) Mittelwerte bekannt

Falls  $\sigma_1 = \sigma_2$  und  $\mu_1, \mu_2$  bekannt, so gilt

$$\frac{S_1^2}{S_2^2} \sim F(n_1, n_2),$$

wobei

$$S_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_1)^2 \quad \text{und} \quad S_2^2 = \frac{1}{n_2} \sum_{j=1}^{n_2} (Y_j - \mu_2)^2.$$

*Testvorschrift* zum Testproblem 1a):

Berechne  $T = \frac{S_1^2}{S_2^2}$  ( $S_1^2$  und  $S_2^2$  wie oben).

Verwirf  $H_0$ , falls  $T > F_{1-\alpha}(n_1, n_2)$ .

Analog für die Testprobleme 1b) und 1c).

*Gütefunktion* beim Testproblem 1a)

Die Gütefunktion ist nur abhängig von  $\lambda = \sigma_1^2 / \sigma_2^2$ :

$$\beta(\lambda) = P \left\{ F(n_1, n_2) > \frac{1}{\lambda} F_{1-\alpha}(n_1, n_2) \right\}.$$

Der Test ist ein UMP- $\alpha$ -Test.

*Bemerkung.*

Das Testproblem 1a) kann verallgemeinert werden zu

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} \leq \lambda_0 \quad \text{gegen} \quad H_1: \frac{\sigma_1^2}{\sigma_2^2} > \lambda_0, \quad \text{wobei } \lambda_0 \text{ bekannt.}$$

(Analog bei 1b) und 1c)). Falls  $\frac{\sigma_1^2}{\sigma_2^2} = \lambda_0$ , so gilt

$$\frac{1}{\lambda_0} \frac{S_1^2}{S_2^2} \sim F(n_1, n_2) \quad (S_1^2 \text{ und } S_2^2 \text{ wie oben}).$$

Weiter wie oben.

Vertrauensgrenzen für  $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$ :

$$\lambda_\ell = T / F_{1-\alpha}(n_1, n_2), \quad \text{wobei} \quad T = \frac{S_1^2}{S_2^2} \quad (S_1^2 \text{ und } S_2^2 \text{ wie oben});$$

$$\lambda_r = T / F_\alpha(n_1, n_2).$$

## b) Mittelwerte unbekannt

Falls  $\sigma_1 = \sigma_2$  und  $\mu_1, \mu_2$  unbekannt, so gilt

$$\frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1),$$

wobei

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2 \quad \text{und} \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2.$$

Alles analog wie bei bekanntem  $\mu_1$  und  $\mu_2$ , jedoch mit  $\bar{X}$  statt  $\mu_1$  und mit  $\bar{Y}$  statt  $\mu_2$ , und mit der F-Verteilung  $F(n_1 - 1, n_2 - 1)$  anstelle von  $F(n_1, n_2)$ . Die einseitigen Tests sind hier UMPU- $\alpha$ -Tests.

## 8. Multinomialverteilung: Einstichprobenproblem

### 1. Zusammenhang zwischen Multinomial- und Chi-Quadrat-Verteilung

*Chi-Quadrat-Verteilung:*

$Z_1, \dots, Z_f$  unabhängig, je  $N(0,1)$ ;

$$Y = Z_1^2 + \dots + Z_f^2;$$

$$\Rightarrow Y \sim \chi^2(f)$$

*Multinomialverteilung:*

$$(X_1, \dots, X_r) \sim \text{Mu}(n, \vartheta_1, \dots, \vartheta_r);$$

$(n_1, \dots, n_r)$  Realisationen zu  $(X_1, \dots, X_r)$ ;

Wir betrachten die Zufallsgröße:

$$T = \sum_{i=1}^r \frac{(X_i - e_i)^2}{e_i} \quad \text{wobei } e_i = E(X_i) = n\vartheta_i;$$

Es gilt:

$$T \xrightarrow[n \rightarrow \infty]{\substack{d \\ (\vartheta_i \text{ fest})}} \chi^2(r-1) \quad (\text{Konvergenz in Verteilung}).$$

Die Zufallsgröße  $T$  besitzt also asymptotisch eine Chi-Quadrat-Verteilung. Die exakte Verteilung der Zufallsgröße  $T$  kann mit Hilfe der Multinomialverteilung berechnet werden.

### 2. Einstichprobenproblem bei der Multinomialverteilung

$$\mathbf{X} = (X_1, \dots, X_r) \sim \text{Mu}(n, \vartheta_1, \dots, \vartheta_r);$$

$n$  bekannt,  $\vartheta_1, \dots, \vartheta_r$  unbekannt;

$$H_0: (\vartheta_1, \dots, \vartheta_r) = (\vartheta_1^0, \dots, \vartheta_r^0) \quad \text{gegen} \quad H_1: (\vartheta_1, \dots, \vartheta_r) \neq (\vartheta_1^0, \dots, \vartheta_r^0);$$

$\mathbf{x} = (n_1, \dots, n_r)$  Realisation zu  $\mathbf{X} = (X_1, \dots, X_r)$ .

Sind die Realisationen  $n_1, \dots, n_r$  vereinbar mit  $H_0$ ?

### 3. Klassischer Chi-Quadrat-Test

Berechne zu gegebenem  $\mathbf{x} = (n_1, \dots, n_r)$  die Testgröße (sogenannte Chi-Quadrat-Testgröße)

$$T(\mathbf{x}) = \sum_{i=1}^r \frac{(n_i - e_i)^2}{e_i} \quad \text{wobei } e_i = E_0(X_i) = n\vartheta_i^0.$$

Unter  $H_0$  besitzt die Testgröße  $T(\mathbf{X})$  (= Testgröße als Zufallsgröße) asymptotisch eine Chi-Quadrat-Verteilung mit  $r-1$  Freiheitsgraden. Daher lautet die Testvorschrift:

$$\text{Verwirf } H_0, \text{ falls } T(\mathbf{x}) > \chi_{1-\alpha}^2(r-1).$$

Übliche Faustregel:  $e_i \geq 5$  für  $i = 1, \dots, r$ .

#### 4. Exakter Chi-Quadrat-Test

Unter  $H_0$  ist die exakte Verteilung der Testgröße

$$T(\mathbf{X}) = \sum_{i=1}^r \frac{(X_i - e_i)^2}{e_i}, \text{ wobei } e_i = E_0(X_i) = n\vartheta_i^0,$$

(= Chi-Quadrat-Testgröße als Zufallsgröße) bekannt; sie kann berechnet werden mit Hilfe der Multinomialverteilung.

*Testvorschrift* zum exakten Chi-Quadrat-Test:

Berechne für die vorliegenden Beobachtungen  $\mathbf{x}_{\text{obs}} = (n_1, \dots, n_r)$  den Wert der Testgröße  $T$ :

$$t_{\text{obs}} = T(\mathbf{x}_{\text{obs}}) = \sum_{i=1}^r \frac{(n_i - e_i)^2}{e_i} \text{ wobei } e_i = n\vartheta_i^0.$$

Berechne dann den exakten  $p$ -Wert

$$p_r = P_0 \{T(\mathbf{X}) \geq t_{\text{obs}}\}.$$

Verwirf  $H_0$ , falls  $p_r \leq \alpha$ .

Praktische Durchführung mit StatXact: Statistics, One-Sample, Chi-Square ...

#### 5. Weitere Testkriterien

Notation:  $\mathbf{X} = (X_1, \dots, X_r)$ ,  $\mathbf{x} = (n_1, \dots, n_r)$ .

a) Chi-Quadrat-Kriterium (wie oben):

$$T(\mathbf{x}) = \sum_{i=1}^r \frac{(n_i - e_i)^2}{e_i} \text{ wobei } e_i = n\vartheta_i^0.$$

Verwirf  $H_0$ , falls  $T(\mathbf{x})$  zu groß.

Asymptotisch:  $T(\mathbf{X}) \sim \chi^2(r-1)$ .

b) LQ-Kriterium (verallgemeinerter Likelihood-Quotienten-Test):

$$T(\mathbf{x}) = \prod_{i=1}^r \left( \frac{n_i}{e_i} \right)^{n_i} \text{ wobei } e_i = n\vartheta_i^0.$$

Verwirf  $H_0$ , falls  $T(\mathbf{x})$  zu groß.

Asymptotisch:  $2 \ln(T(\mathbf{X})) \sim \chi^2(r-1)$ .

c) Wahrscheinlichkeits-Kriterium von Freeman-Halton:

$$T(\mathbf{x}) = P_0 \{ \mathbf{X} = \mathbf{x} \} = P_0 \{ X_1 = n_1, \dots, X_r = n_r \} = \frac{n!}{n_1! \dots n_r!} (\vartheta_1^0)^{n_1} \dots (\vartheta_r^0)^{n_r}.$$

Verwirf  $H_0$ , falls  $T(\mathbf{x})$  zu klein.

Asymptotisch:  $-2 \ln(T(\mathbf{X})) \sim \chi^2(r-1)$ .

Analog zum exakten Chi-Quadrat-Test können auch mit den Statistiken b) und c) exakte Tests zu unserem Testproblem durchgeführt werden.

## 6. Monte-Carlo-Variante

$t_{\text{obs}} = T(\mathbf{x}_{\text{obs}})$  = beobachteter Wert der Teststatistik;

$p_r = P_0\{T(\mathbf{X}) \geq t_{\text{obs}}\}$  = rechtsseitiger  $p$ -Wert.

Bestimmung des  $p$ -Werts  $p_r$  durch Simulation:

- Erzeuge  $M$  zufällige Realisationen von  $\mathbf{X} = (X_1, \dots, X_r)$  unter  $H_0: \mathbf{x}_1, \dots, \mathbf{x}_M$ . Jedes  $\mathbf{x}_i$  ist ein Vektor  $(n_1, \dots, n_r)$ .
- Bestimme den Wert der Zufallsgröße  
 $Y = \#\{i \mid T(\mathbf{x}_i) \geq t_{\text{obs}}\}$  = Anzahl der Realisationen  $\mathbf{x}_i$  mit  $T(\mathbf{x}_i) \geq t_{\text{obs}}$ .  
 Es gilt:  $Y \sim \text{Bi}(M, p_r)$  und somit:  $Y/M \approx p_r$ .  
 Vertrauensgrenzen für  $p_r$  mit den bekannten Verfahren für die Binomialverteilung.

## 7. Poissonmodell

$X_1, \dots, X_r$  unabhängige Zufallsgrößen, wobei  $X_i \sim \text{Po}(\lambda_i)$ ,  $i = 1, \dots, r$ ;

$H_0: \lambda_1 = \dots = \lambda_r$  gegen  $H_1$ : nicht alle  $\lambda_i$  gleich.

Es sei  $S = X_1 + \dots + X_r$ . Dann gilt:

$$(X_1, \dots, X_r) | S=n \sim \text{Mu}(n, \vartheta_1, \dots, \vartheta_r) \text{ wobei } \vartheta_i = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_r}$$

und weiter

$$\lambda_1 = \dots = \lambda_r \Leftrightarrow \vartheta_1 = \dots = \vartheta_r = 1/r.$$

Damit ist das Testproblem reduziert auf das Einstichprobenproblem bei der Multinomialverteilung.

## 8. Chi-Quadrat-Anpassungstest (Goodness of fit test)

*Beispiel:*

$Y_1, \dots, Y_n$  unabhängig, je mit gleicher Verteilungsfunktion

$H_0: Y_j \sim N(\mu, \sigma^2)$ ,  $\mu$  und  $\sigma^2$  unbekannt;

$H_1: Y_j$  nicht normalverteilt.

Üblicher Lösungsvorschlag:

- Wertebereich der  $Y_i$  in  $r$  Teilintervalle zerlegen:  $I_1, \dots, I_r$ ;

$$\left. \begin{array}{l} X_i = \#\{j \mid Y_j \in I_i\} \\ \vartheta_i = P\{Y_j \in I_i\} \end{array} \right\} \Rightarrow (X_1, \dots, X_r) \sim \text{Mu}(n, \vartheta_1, \dots, \vartheta_r).$$

Unter  $H_0$ :  $\vartheta_i = \vartheta_i^0 = P_0\{Y_j \in I_i\}$  (abhängig von unbekanntem  $\mu$  und  $\sigma^2$ );

- Unter  $H_0$   $\mu$  und  $\sigma^2$  (asymptotisch effizient) schätzen mit vorliegenden Daten  
 $\rightarrow \hat{\vartheta}_1^0, \dots, \hat{\vartheta}_r^0$ ;
- Klassischen Chi-Quadrat-Test zum Einstichprobenproblem mit geschätzten Parametern  $\hat{\vartheta}_1^0, \dots, \hat{\vartheta}_r^0$  durchführen. Anzahl der Freiheitsgrade:  $f = r - s - 1$ , wobei  $r$  = Anzahl der Klassen und  $s$  = Anzahl der geschätzten Modellparameter. Im vorliegenden Beispiel: Zwei Modellparameter  $\mu$  und  $\sigma^2$ ; daher  $f = r - s - 1 = r - 2 - 1 = r - 3$ .

## 9. Multinomialverteilung: Mehrstichprobenproblem

### 1. Problemstellung

$$\mathbf{X}_1 = (X_{11}, \dots, X_{1s}) \sim \text{Mu}(n_1, \vartheta_{11}, \dots, \vartheta_{1s});$$

$$\vdots$$

$$\mathbf{X}_r = (X_{r1}, \dots, X_{rs}) \sim \text{Mu}(n_r, \vartheta_{r1}, \dots, \vartheta_{rs});$$

$\mathbf{X}_1, \dots, \mathbf{X}_r$  unabhängige Zufallsvektoren;

$n_1, \dots, n_r$  bekannt,  $\vartheta_{ij}$  unbekannt ( $i=1, \dots, r; j=1, \dots, s$ );

$H_0: \vartheta_{ij} = c_j \quad \forall i, j$  (d.h. in allen  $r$  Stichproben gleiche Parameter);

$H_1$ : Negation zu  $H_0$ ;

$\mathbf{x}_i = (n_{i1}, \dots, n_{is})$  Realisation zu  $\mathbf{X}_i = (X_{i1}, \dots, X_{is})$ .

Sind die Realisationen  $n_{ij}$  vereinbar mit  $H_0$ ?

In Tabellenform:

Zufallsgrößen					Parameter					Realisationen				
	1	...	s			1	...	s			1	...	s	
$\mathbf{X}_1$	$X_{11}$	...	$X_{1s}$	$n_1$	$\mathbf{X}_1$	$\vartheta_{11}$	...	$\vartheta_{1s}$	1	$\mathbf{X}_1$	$n_{11}$	...	$n_{1s}$	$n_1$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$\mathbf{X}_r$	$X_{r1}$	...	$X_{rs}$	$n_r$	$\mathbf{X}_r$	$\vartheta_{r1}$	...	$\vartheta_{rs}$	1	$\mathbf{X}_r$	$n_{r1}$	...	$n_{rs}$	$n_r$
	$X_{\bullet 1}$	...	$X_{\bullet s}$	$n$							$n_{\bullet 1}$	...	$n_{\bullet s}$	$n$

### 2. Klassischer Chi-Quadrat-Test

Notation:  $\mathbf{X} = (X_{ij}) = (r \times s)$ ,  $\mathbf{x} = (n_{ij}) = (r \times s)$ ,  $n_{i\bullet} = n_i$  für  $i=1, \dots, r$ .

Berechne zu gegebenem  $\mathbf{x} = (n_{ij})$  die Testgröße (sogenannte Chi-Quadrat-Testgröße)

$$T(\mathbf{x}) = \sum_{i,j=1}^{r,s} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad \text{wobei} \quad e_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}.$$

Unter  $H_0$  besitzt die Testgröße  $T(\mathbf{X})$  (= Testgröße als Zufallsgröße) asymptotisch (für  $n \rightarrow \infty$ ) eine Chi-Quadrat-Verteilung mit  $f = (r-1)(s-1)$  Freiheitsgraden. Daher lautet die Testvorschrift:

$$\text{Verwirf } H_0, \text{ falls } T(\mathbf{x}) \geq \chi_{1-\alpha}^2(f), \text{ wobei } f = (r-1)(s-1).$$

Übliche Faustregel:  $e_{ij} \geq 5 \quad \forall i, j$ .

### 3. Exakter Chi-Quadrat-Test

Es gilt:

$$P_0 \{ \mathbf{X} = \mathbf{x} \mid X_{\bullet 1} = n_{\bullet 1}, \dots, X_{\bullet s} = n_{\bullet s} \} = \frac{\prod_{i=1}^r n_{i\bullet}! \prod_{j=1}^s n_{\bullet j}!}{n! \prod_{i,j=1}^{r,s} n_{ij}!}.$$

Für  $r = s = 2$  ist dies die gewöhnliche hypergeometrische Verteilung, die wir ja beim exakten Vierfeldertest angetroffen haben. Die obige Verteilung ist also eine Verallgemeinerung der hypergeometrischen Verteilung auf den multivariaten (mehrdimensionalen) Fall, da ja

$\mathbf{X} = (X_{ij}) = (r \times s)$  eine Matrix ist. Die Verteilung ist festgelegt durch die Zeilen- und Spaltensummen  $n_{i\cdot}$  und  $n_{\cdot j}$  ( $i = 1, \dots, r; j = 1, \dots, s$ ). Wir bezeichnen diese Verteilung mit  $H(n_{i\cdot}, n_{\cdot j}; i, j = 1, \dots, r, s)$ .

*Testvorschrift* zum exakten Chi-Quadrat-Test:

Berechne für die vorliegenden Beobachtungen  $\mathbf{x}_{\text{obs}} = (n_{ij}) = (r \times s)$  den Wert der Testgröße  $T$ :

$$t_{\text{obs}} = T(\mathbf{x}_{\text{obs}}) = \sum_{i,j=1}^{r,s} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad \text{wobei} \quad e_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}.$$

Berechne dann den exakten  $p$ -Wert

$$p_r = P_0 \{T(\mathbf{X}) \geq t_{\text{obs}}\}.$$

Verwirf  $H_0$ , falls  $p_r \leq \alpha$ .

Praktische Durchführung mit StatXact:

Statistics, Unordered  $r \times c$  tables, Pearson's Chi-Square, ...

#### 4. Einfaches Modell für die verallgemeinerte hypergeometrische Verteilung

Ein Kartendeck enthält  $n$  Spielkarten in  $r$  Farben:

$n_{1\cdot}$  Karten der Farbe  $F_1$ ;

...

$n_{r\cdot}$  Karten der Farbe  $F_r$ .

	$S_1$	...	$S_s$	
$F_1$		...		$n_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$F_r$		...		$n_{r\cdot}$
	$n_{\cdot 1}$	...	$n_{\cdot s}$	$n$

Die Karten werden gemischt und zufällig an die  $s$  Spieler verteilt:

$n_{\cdot 1}$  Karten an Spieler  $S_1$ ;

...

$n_{\cdot s}$  Karten an Spieler  $S_s$ .

$Y_{ij}$  = Anzahl der Karten der Farbe  $F_i$  beim Spieler  $S_j$ ;

$\mathbf{Y} = (Y_{ij}) = (r \times s)$  Matrix aller Zufallsgrößen  $Y_{ij}$ ;

$\mathbf{y} = (n_{ij}) = (r \times s)$  Realisation der zufälligen Matrix  $\mathbf{Y}$ .

$$\text{Nun gilt: } P\{\mathbf{Y} = \mathbf{y}\} = \frac{\prod_{i=1}^r n_{i\cdot}! \prod_{j=1}^s n_{\cdot j}!}{n! \prod_{i,j=1}^{r,s} n_{ij}!}, \quad \text{d.h. } \mathbf{Y} \sim H(n_{i\cdot}, n_{\cdot j}; i, j = 1, \dots, r, s).$$

Weiter gilt:

$$Y_{ij} \sim H(n, n_{i\cdot}, n_{\cdot j}) \quad \forall i, j;$$

$$E(Y_{ij}) = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n} = e_{ij} \quad \forall i, j.$$

## 5. Weitere Testkriterien

Notation:  $\mathbf{X} = (X_{ij}) = (r \times s)$ ,  $\mathbf{x} = (n_{ij}) = (r \times s)$ .

a) Chi-Quadrat-Kriterium (wie oben):

$$T(\mathbf{x}) = \sum_{i,j=1}^{r,s} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad \text{wobei} \quad e_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}.$$

Verwirf  $H_0$ , falls  $T(\mathbf{x})$  zu groß.

Asymptotisch:  $T(\mathbf{X}) \sim \chi^2(f)$  wobei  $f = (r-1)(s-1)$ .

b) LQ-Kriterium (verallgemeinerter Likelihood-Quotienten-Test):

$$T(\mathbf{x}) = \prod_{i,j=1}^{r,s} \left( \frac{n_{ij}}{e_{ij}} \right)^{n_{ij}} \quad (e_{ij} \text{ wie oben}).$$

Verwirf  $H_0$ , falls  $T(\mathbf{x})$  zu groß.

Asymptotisch:  $2 \ln(T(\mathbf{X})) \sim \chi^2(f)$  wobei  $f = (r-1)(s-1)$ .

c) Wahrscheinlichkeits-Kriterium von Freeman-Halton:

$$T(\mathbf{x}) = P_0 \{ \mathbf{X} = \mathbf{x} \mid X_{\cdot 1} = n_{\cdot 1}, \dots, X_{\cdot s} = n_{\cdot s} \} = \frac{\prod_{i=1}^r n_{i\cdot}! \prod_{j=1}^s n_{\cdot j}!}{n! \prod_{i,j=1}^{r,s} n_{ij}!}$$

Verwirf  $H_0$ , falls  $T(\mathbf{x})$  zu klein.

Asymptotisch:  $-2 \ln(T(\mathbf{X})) \sim \chi^2(f)$  wobei  $f = (r-1)(s-1)$ .

Mit StatXact können analog zum exakten Chi-Quadrat-Test auch zu den Statistiken b) und c) exakte Tests zu unserem Testproblem durchgeführt werden:

Statistics, Unordered  $r \times c$  tables,  $\left\{ \begin{array}{l} (a) \text{ Pearson's Chi-Square, ...} \\ (b) \text{ Likelihood Ratio, ...} \\ (c) \text{ Fisher-Freeman-Halton, ...} \end{array} \right.$

## 6. Monte-Carlo-Variante

$\mathbf{x}_{\text{obs}} = (n_{ij}) = (r \times s)$  = beobachtete Häufigkeitstabelle mit Randsummen  $n_{i\cdot}$  ( $= n_{i\cdot}$ ) und  $n_{\cdot j}$ ;

$t_{\text{obs}} = T(\mathbf{x}_{\text{obs}})$  = beobachteter Wert der Teststatistik;

$p_r = P_0 \{ T(\mathbf{X}) \geq t_{\text{obs}} \mid X_{\cdot 1} = n_{\cdot 1}, \dots, X_{\cdot s} = n_{\cdot s} \}$  = rechtsseitiger  $p$ -Wert.

Bestimmung des  $p$ -Werts  $p_r$  durch Simulation:

- Erzeuge  $M$  zufällige Realisationen von  $\mathbf{X} = (X_{ij}) = (r \times s)$  unter  $H_0$  bei gegebenen Spaltensummen  $X_j = n_j$  d.h. gemäß der verallgemeinerten hypergeometrischen Verteilung  $H(n_{i\cdot}, n_{\cdot j}; i, j = 1, \dots, r, s)$ :  $\mathbf{x}_1 \dots, \mathbf{x}_M$ . Jedes  $\mathbf{x}_i$  ist eine Matrix  $(n_{ij}) = (r \times s)$  mit den vorgegebenen Randsummen  $n_{i\cdot}, n_{\cdot j}$  ( $i, j = 1, \dots, r, s$ ).

- Bestimme den Wert der Zufallsgröße

$$Y = \# \{ i \mid T(\mathbf{x}_i) \geq t_{\text{obs}} \} = \text{Anzahl der Realisationen } \mathbf{x}_i \text{ mit } T(\mathbf{x}_i) \geq t_{\text{obs}}.$$

Es gilt:  $Y \sim \text{Bi}(M, p_r)$  und somit:  $Y/M \approx p_r$ .

Vertrauensgrenzen für  $p_r$  mit den bekannten Verfahren für die Binomialverteilung.

## 7. Kontingenzmodell

Bisher: Mehrstichprobenmodell bei der Multinomialverteilung

$$\mathbf{X}_i = (X_{i1}, \dots, X_{is}) \sim \text{Mu}(n_i, \vartheta_{i1}, \dots, \vartheta_{is}) \quad \text{für } i = 1, \dots, r;$$

$\mathbf{X}_1, \dots, \mathbf{X}_r$  unabhängige Zufallsvektoren.

In Tabellenform:

	1	...	s	
$\mathbf{X}_1$	$X_{11}$	...	$X_{1s}$	$n_1$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\mathbf{X}_r$	$X_{r1}$	...	$X_{rs}$	$n_r$
	$X_{\bullet 1}$	...	$X_{\bullet s}$	$n$

Feste Zeilensummen  $n_1, \dots, n_r$ ;  
zufällige Spaltensummen  $X_{\bullet 1}, \dots, X_{\bullet s}$ .

Jetzt: Kontingenzmodell

$$\mathbf{X} = (X_{ij}) = (r \times s) \sim \text{Mu}(n, \vartheta) \quad \text{mit } \vartheta = (\vartheta_{ij}) = (r \times s)$$

In Tabellenform:

Zufallsgrößen					Parameter					Realisationen							
A	B	1	...	s		A	B	1	...	s		A	B	1	...	s	
1		$X_{11}$	...	$X_{1s}$	$X_{1\bullet}$	1		$\vartheta_{11}$	...	$\vartheta_{1s}$	$\vartheta_{1\bullet}$	1		$n_{11}$	...	$n_{1s}$	$n_{1\bullet}$
$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$
r		$X_{r1}$	...	$X_{rs}$	$X_{r\bullet}$	r		$\vartheta_{r1}$	...	$\vartheta_{rs}$	$\vartheta_{r\bullet}$	r		$n_{r1}$	...	$n_{rs}$	$n_{r\bullet}$
		$X_{\bullet 1}$	...	$X_{\bullet s}$	$n$			$\vartheta_{\bullet 1}$	...	$\vartheta_{\bullet s}$	1			$n_{\bullet 1}$	...	$n_{\bullet s}$	$n$

Hier: Zufällige Zeilensummen  $X_{1\bullet}, \dots, X_{r\bullet}$  ;  
zufällige Spaltensummen  $X_{\bullet 1}, \dots, X_{\bullet s}$  ;  
feste Gesamtsumme  $n$ .

$$H_0 : \vartheta_{ij} = \vartheta_{i\bullet} \vartheta_{\bullet j} \quad \forall i, j \quad (\text{d.h. die Merkmale A und B sind unabhängig});$$

$H_1$  : Negation zu  $H_0$ ;

$$\mathbf{x} = (n_{ij}) = (r \times s) \text{ Realisation zu } \mathbf{X} = (X_{ij}) = (r \times s).$$

Sind die Realisationen  $n_{ij}$  vereinbar mit  $H_0$  ?

Das bedingte Problem bei gegebenen Zeilensummen ist äquivalent zum Mehrstichprobenproblem bei der Multinomialverteilung. Daher kann man die gleichen asymptotischen und exakten Verfahren anwenden wie oben.

## 8. Poissonmodell

$X_{ij} \sim \text{Po}(\lambda_{ij})$  für  $i = 1, \dots, r$  und  $j = 1, \dots, s$ ;

alle  $X_{ij}$  stochastisch unabhängig.

In Tabellenform:

Zufallsgrößen					Parameter					Realisationen							
A \ B		1	...	s		A \ B		1	...	s		A \ B		1	...	s	
1		$X_{11}$	...	$X_{1s}$	$X_{1\bullet}$	1		$\lambda_{11}$	...	$\lambda_{1s}$	$\lambda_{1\bullet}$	1		$n_{11}$	...	$n_{1s}$	$n_{1\bullet}$
$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$
r		$X_{r1}$	...	$X_{rs}$	$X_{r\bullet}$	r		$\lambda_{r1}$	...	$\lambda_{rs}$	$\lambda_{r\bullet}$	r		$n_{r1}$	...	$n_{rs}$	$n_{r\bullet}$
		$X_{\bullet 1}$	...	$X_{\bullet s}$	$X_{\bullet\bullet}$			$\lambda_{\bullet 1}$	...	$\lambda_{\bullet s}$	$\lambda_{\bullet\bullet}$			$n_{\bullet 1}$	...	$n_{\bullet s}$	$n_{\bullet\bullet} (= n)$

Hier: Zufällige Zeilensummen  $X_{1\bullet}, \dots, X_{r\bullet}$ ;  
 zufällige Spaltensummen  $X_{\bullet 1}, \dots, X_{\bullet s}$ ;  
 zufällige Gesamtsumme  $X_{\bullet\bullet}$ .

$$H_0: \lambda_{ij} = \frac{\lambda_{i\bullet} \lambda_{\bullet j}}{\lambda_{\bullet\bullet}} \quad \forall i, j \quad (\text{d.h. die Merkmale A und B sind unabhängig});$$

$H_1$ : Negation zu  $H_0$ ;

$\mathbf{x} = (n_{ij}) = (r \times s)$  Realisation zu  $\mathbf{X} = (X_{ij}) = (r \times s)$ .

Sind die Realisationen  $n_{ij}$  vereinbar mit  $H_0$ ?

Das bedingte Problem bei gegebener Gesamtsumme  $X = n$  ist äquivalent zum Kontingenzproblem bei der Multinomialverteilung. Daher kann man die gleichen asymptotischen und exakten Verfahren anwenden wie oben.

## 10. Vorzeichentest und Test von McNemar

### 1. Vorzeichentest

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

$$P\{X_i = +1\} = p_+ \quad (\text{positive Wirkung});$$

$$P\{X_i = -1\} = p_- \quad (\text{negative Wirkung});$$

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

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

$$H_0: p_+ \leq p_- \quad \text{gegen} \quad H_1: p_+ > p_-.$$

Es sei

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

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

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

Dann gilt:

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

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

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

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

*Vorzeichentest:*

Berechne  $m = N_+ + N_-$ .

Bestimme  $k_r =$  rechte kritische  $\alpha$ -Schranke zu  $\text{Bi}(m, \frac{1}{2})$ .

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

Der Test hält das Signifikanzniveau  $\alpha$  ein, und er ist konsistent.

(Bei der zweiseitigen und der anderen einseitigen Fragestellung ist alles analog.)

### 2. Test von McNemar (Symmetriepfung in einer Vier-Felder-Tafel):

Zwei typische Anwendungen:

a)  $n$  Versuchspersonen sollen zwei Aufgaben A und B lösen;

Gesamtheit.

	A \ B	+	-	
A	B	$p_{11}$	$p_{12}$	
+		$p_{21}$	$p_{22}$	
-				1

Stichprobe:

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

$p_{11} =$  Wahrscheinlichkeit, daß eine (zufällig ausgewählte) Vp. A und B löst;

...

$X_{11} =$  Anzahl der Vp., welche A und B lösen;

...

- b)  $n$  Versuchspersonen werden zweimal befragt bzw. untersucht  
(A=vorher, B=nachher);

Gesamtheit.

	B		
A	+	-	
+	$p_{11}$	$p_{12}$	
-	$p_{21}$	$p_{22}$	
			1

Stichprobe:

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

$p_{11}$  = Wahrscheinlichkeit, daß eine (zufällig ausgewählte) Vp.  
vorher und nachher "positiv" ist;

...

$X_{11}$  = Anzahl der Vp., welche vorher und nachher "positiv" sind;

...

Für beide Anwendungen gilt das folgende Modell:

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

$$H_0 : p_{12} \leq p_{21} \quad \text{gegen} \quad H_1 : p_{12} > p_{21}$$

Beachte:

$$p_{12} = p_{21} \Leftrightarrow \text{Matrix } (p_{ij}) \text{ ist symmetrisch.}$$

Daher die Bezeichnung *Symmetrieprüfung in einer Kontingenztafel*.

Es gilt:

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

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

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

*Test von McNemar:*

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

Bestimme  $k_r$  = rechte kritische  $\alpha$ -Schranke zu  $\text{Bi}(m, \frac{1}{2})$ .

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

Der Test hält das Signifikanzniveau  $\alpha$  ein, und er ist konsistent.

(Bei der zweiseitigen und der anderen einseitigen Fragestellung ist alles analog.)

## 11. Asymptotische relative Effizienz zwischen t-Test und Vorzeichentest

### 1. t-Test und Vorzeichentest

$X_1, \dots, X_n$  unabhängige Zufallsgrößen je mit der gleichen Dichte  $f_\mu$ ;

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

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

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

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

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

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

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  $n \rightarrow \infty$  interessieren, so können auch andere Verteilungen in Betracht gezogen werden.

*Vorzeichentest:*

$$N_+ = \#\{i \mid X_i > 0\};$$

$$N_- = \#\{i \mid X_i < 0\};$$

(da  $F$  stetig:  $P\{X_i = 0\} = 0$ ).

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

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

### 2. Asymptotische Gütefunktion des t-Tests

Gütefunktion des t-Tests:

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

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  $n \rightarrow \infty$  ( $\delta$  und  $\alpha$  fest) gilt  $S \xrightarrow{P} \sigma$  und

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

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

### 3. Asymptotische Gütefunktion des Vorzeichentests

Gütefunktion des Vorzeichentests:

$$\beta_n(\mu) = \mathbb{P}_\mu \{N_+ > k_r(n, \alpha)\};$$

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

$$N_+ \sim \text{Bi}(n, \vartheta),$$

wobei

$$\begin{aligned} \vartheta &= \mathbb{P}\{X_i > \mu_0\} = 1 - F_\mu(\mu_0) = 1 - F(\mu_0 - \mu) = 1 - F(-\delta/\sqrt{n}) \\ &= 1 - F(0) + \frac{\delta}{\sqrt{n}} f(0) + O(1/n) = \frac{1}{2} + \frac{\delta}{\sqrt{n}} f(0) + O(1/n); \\ \sqrt{n}(\vartheta - \frac{1}{2}) &\xrightarrow[n \rightarrow \infty]{\delta \text{ fest}} \delta f(0). \end{aligned}$$

Für die Gütefunktion im  $\delta$ -Maßstab gilt bei  $n \rightarrow \infty$  ( $\delta$  und  $\alpha$  fest):

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

Nun gilt

$$\begin{aligned} \frac{N_+ - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} &\xrightarrow{d} N(0,1) \\ \frac{k_r(n, \alpha) - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} &= \frac{k_r(n, \alpha) - n/2}{\sqrt{n}/2} \cdot \frac{\sqrt{n}/2}{\sqrt{n\vartheta(1-\vartheta)}} + \frac{n/2 - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \\ &= \frac{k_r(n, \alpha) - n/2}{\sqrt{n}/2} \cdot \frac{1}{2\sqrt{\vartheta(1-\vartheta)}} - \frac{\sqrt{n}(\vartheta - \frac{1}{2})}{\sqrt{\vartheta(1-\vartheta)}} \rightarrow z_{1-\alpha} - 2\delta f(0), \end{aligned}$$

$\xrightarrow{\rightarrow z_{1-\alpha}} \quad \xrightarrow{\rightarrow 1} \quad \xrightarrow{\rightarrow 2\delta f(0)}$

und somit

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

### 4. Asymptotische relative Effizienz

Es sei

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

$n_B$  Stichprobenumfang beim Vorzeichentest (Verfahren B);

$$\lambda_n = \frac{n_B}{n} \left( = \frac{n_B}{n_A} \right).$$

Wir betrachten nun den Grenzübergang  $n$  ( $= n_A$ )  $\rightarrow \infty$ ,  $\lambda_n \rightarrow \lambda$ , und wählen dabei  $\lambda$  so, daß die beiden asymptotischen Gütefunktionen übereinstimmen (bei entsprechenden  $\delta$ -Werten). Die

asymptotische relative Effizienz (are = asymptotic relative efficiency) zwischen dem t-Test und dem Vorzeichentest wird dann definiert als:

$$\text{are}(A : B) = \lim \frac{n_B}{n_A} = \lim \lambda_n = \lambda.$$

Wir haben oben die asymptotischen Gütefunktionen bestimmt:

$$\beta_n^A(\delta) \longrightarrow \mathbb{P}\{N(0,1) < -z_{1-\alpha} + \delta/\sigma\},$$

$$\beta_{n_B}^B(\delta) \longrightarrow \mathbb{P}\{N(0,1) < -z_{1-\alpha} + 2\delta f(0)\}.$$

Dabei ist

$$\text{beim Grenzprozess A: } \delta = \delta_A = \sqrt{n_A}(\mu - \mu_0);$$

$$\text{beim Grenzprozess B: } \delta = \delta_B = \sqrt{n_B}(\mu - \mu_0) = \sqrt{n_B/n_A} \cdot \sqrt{n_A}(\mu - \mu_0) = \sqrt{\lambda_n} \delta_A$$

Um die asymptotischen Gütefunktionen vergleichen zu können, muß sich der Parameter  $\delta$  bei beiden Grenzprozessen auf die gleiche Alternative  $d = \mu - \mu_0$  beziehen; wir setzen daher

$$\text{beim Grenzprozess A: } \delta_A = \delta = \sqrt{n}(\mu - \mu_0) \quad (\text{beachte, daß } n = n_A),$$

$$\text{beim Grenzprozess B: } \delta_B = \sqrt{\lambda_n} \delta = \sqrt{n_B}(\mu - \mu_0).$$

Da  $\lambda_n \rightarrow \lambda$  konvergiert, so erhalten wir

$$\beta_{n_B}^B(\delta\sqrt{\lambda_n}) \longrightarrow \mathbb{P}\{N(0,1) < -z_{1-\alpha} + 2\delta\sqrt{\lambda} f(0)\}.$$

Die beiden asymptotischen Gütefunktionen zu A und B sind identisch, falls

$$2\sqrt{\lambda} f(0) = \frac{1}{\sigma}, \quad \text{d.h. falls } \lambda = \frac{1}{4\sigma^2 f^2(0)}.$$

Somit beträgt die asymptotische relative Effizienz des t-Tests (Verfahren A) im Vergleich zum Zeichentest (Verfahren B)

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

*Bemerkung:*

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

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

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

Die dort gegebene Tabelle 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  $\text{are}(A : B) = \pi/2 = 1.57$ . Bei anderen Verteilungen kann der Zeichentest effizienter sein als der t-Test.

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

### 1. Problemstellung

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

A \ B	+	-	
+	$p_{11}$	$p_{12}$	
-	$p_{21}$	$p_{22}$	
			1

$p_{11}$  = W'keit, daß eine (zufällig ausgewählte) Vp. vorher und nachher "positiv" ist

$p_{12}$  = W'keit, daß eine (zufällig ausgewählte) Vp. vorher "positiv" und nachher "negativ" ist

...

$H_0: p_{12} \leq p_{21}$  gegen  $H_1: p_{12} > p_{21}$

(Bei der zweiseitigen und der anderen einseitigen 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 Vp., welche vorher "positiv" sind ;

$X_2$  = Anzahl der Vp., welche nachher "positiv" sind ;

Modell:

$$X_1 \sim \text{Bi}(n, \vartheta_1), \quad \vartheta_1 = p_{11} + p_{12};$$

$$X_2 \sim \text{Bi}(n, \vartheta_2), \quad \vartheta_2 = p_{11} + p_{21};$$

$X_1, X_2$  unabhängig.

$$H_0: p_{12} \leq p_{21} \quad \Leftrightarrow \quad \vartheta_1 \leq \vartheta_2$$

$$H_1: p_{12} > p_{21} \quad \Leftrightarrow \quad \vartheta_1 > \vartheta_2.$$

	+	-	
$X_1$	$k_1$	$\cdot$	$n$
$X_2$	$k_2$	$\cdot$	$n$
$X$	$k$	$\cdot$	$2n$

Exakter Vier-Felder-Test von Fisher:

Falls  $\vartheta_1 = \vartheta_2$  d.h. falls  $p_{12} = p_{21}$ , so gilt:

$$X_1 | X = k \sim H(2n, n, k).$$

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

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

Asymptotische Gütefunktion:

ursprüngliche Variablen:  $n, \vartheta_1, \vartheta_2, \alpha$ .

neue Variablen anstelle von  $\vartheta_1$  und  $\vartheta_2$ :

$$\delta = \sqrt{n}(\vartheta_1 - \vartheta_2) = \sqrt{n}(p_{12} - p_{21})$$

$$\bar{\vartheta} = \frac{1}{2}(\vartheta_1 + \vartheta_2) = p_{11} + \frac{1}{2}(p_{12} + p_{21}).$$

Gütefunktion abhängig von  $n, \delta, \bar{\vartheta}, \alpha$ . Für  $n \rightarrow \infty$  ( $\delta, \bar{\vartheta}, \alpha$  fest) gilt:

$$\beta_n^U(\delta) \longrightarrow \mathbf{P} \left\{ \mathbf{N}(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{2\bar{\vartheta}(1-\bar{\vartheta})}} \right\}.$$

(Vgl. Kapitel 4: Asymptotische Gütefunktion zum Vier-Felder-Test; beachte die unterschiedliche Definition von  $n$ .)

### 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);

$X_{11}$  = Anzahl der Vp., welche vorher und nacher "positiv" sind

$X_{12}$  = Anzahl der Vp., welche vorher positiv und nacher "negativ" sind

...

Stichprobe:

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

Modell:

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

$$H_0: p_{12} \leq p_{21} \quad \text{gegen} \quad H_1: p_{12} > p_{21}.$$

Test von McNemar:

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

$$X_{12} \mid X_{12} + X_{21} = m \sim \text{Bi}\left(m, \frac{1}{2}\right).$$

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

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

Asymptotische Gütefunktion:

ursprüngliche Variablen:  $n, p_{12}, p_{21}, \alpha$ .

neue Variablen anstelle von  $p_{12}$  und  $p_{21}$ :

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

$$\bar{p} = \frac{1}{2}(p_{12} + p_{21}).$$

Es gilt dann

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

Die Gütefunktion ist abhängig von  $\otimes = (n, \delta, \bar{p}, \alpha)$ , und wir können schreiben

$$\beta_n^V(\delta) = P_{\otimes} \{ \text{Test von McNemar verwirft } H_0 \}$$

$$(1) \quad = \sum_{m=0}^n P_{\otimes} \{ X_{12} + X_{21} = m \} \underbrace{P_{\otimes} \{ \text{Test verwirft } H_0 \mid X_{12} + X_{21} = m \}}_{= \beta_{n|m}(\delta)}$$

Nun ist  $X_{12} + X_{21} \sim \text{Bi}(n, p_{12} + p_{21})$ , und daher gilt

$$\frac{X_{12} + X_{21}}{n} \xrightarrow{P} p_{12} + p_{21} = 2\bar{p} \quad \text{für } n \rightarrow \infty.$$

Wir betrachten daher bei der bedingten Gütefunktion  $\beta_{n|m}(\delta)$  den Grenzübergang

$$(2) \quad n \rightarrow \infty, m/n \rightarrow 2\bar{p} \quad (\delta, \bar{p}, \alpha \text{ fest}).$$

Es gilt dann

$$\beta_{n|m}(\delta) = P_{\otimes} \{ \text{Test verwirft } H_0 \mid X_{12} + X_{21} = m \} = P_{\otimes} \{ X_{12} > k_r(m, \alpha) \mid X_{12} + X_{21} = m \}$$

Nun ist

$$X_{12} \mid X_{12} + X_{21} = m \sim \text{Bi}(m, \eta),$$

wobei

$$\eta = \frac{p_{12}}{p_{12} + p_{21}} = \frac{1}{2\bar{p}} \left( \bar{p} + \frac{\delta}{2\sqrt{n}} \right) = \frac{1}{2} + \frac{\delta}{4\bar{p}\sqrt{n}},$$

und daher können wir schreiben

$$\begin{aligned} \beta_{\eta|m}(\delta) &= P_{\otimes} \{X_{12} > k_r(m, \alpha) | X_{12} + X_{21} = m\} = P \{ \text{Bi}(m, \eta) > k_r(m, \alpha) \} \\ (3) \quad &= P \left\{ \frac{\text{Bi}(m, \eta) - m\eta}{\sqrt{m\eta(1-\eta)}} > \frac{k_r(m, \alpha) - m\eta}{\sqrt{m\eta(1-\eta)}} \right\}. \end{aligned}$$

Nun gilt

$$\begin{aligned} \frac{\text{Bi}(m, \eta) - m\eta}{\sqrt{m\eta(1-\eta)}} &\xrightarrow{d} N(0,1) \\ (4) \quad \frac{k_r(m, \alpha) - m\eta}{\sqrt{m\eta(1-\eta)}} &= \frac{k_r(m, \alpha) - m/2}{\sqrt{m}/2} \cdot \frac{\sqrt{m}/2}{\sqrt{m\eta(1-\eta)}} + \frac{m(\frac{1}{2} - \eta)}{\sqrt{m\eta(1-\eta)}} \\ &= \underbrace{\frac{k_r(m, \alpha) - m/2}{\sqrt{m}/4}}_{\rightarrow z_{1-\alpha}} \cdot \underbrace{\frac{1}{2\sqrt{\eta(1-\eta)}}}_{\rightarrow 1} - \frac{\sqrt{m}(\eta - \frac{1}{2})}{\sqrt{\eta(1-\eta)}} \end{aligned}$$

Der letzte Term in (4) kann geschrieben werden als

$$\frac{\sqrt{m}(\frac{1}{2} - \eta)}{\sqrt{\eta(1-\eta)}} = \sqrt{m/n} \frac{\sqrt{n}(\frac{1}{2} - \eta)}{\sqrt{\eta(1-\eta)}} \longrightarrow \sqrt{2\bar{p}} \cdot \frac{\delta}{2\bar{p}} = \frac{\delta}{\sqrt{2\bar{p}}},$$

und somit gilt

$$\frac{k_r(m, \alpha) - m\eta}{\sqrt{m\eta(1-\eta)}} \longrightarrow z_{1-\alpha} - \frac{\delta}{\sqrt{2\bar{p}}},$$

und wir erhalten aus (3)

$$\beta_{\eta|m}(\delta) \longrightarrow P \left\{ N(0,1) > z_{1-\alpha} - \frac{\delta}{\sqrt{2\bar{p}}} \right\} = P \left\{ N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{2\bar{p}}} \right\}.$$

Und nun folgt (wie beim Vier-Felder-Test), daß auch die unbedingte Gütefunktion den gleichen Grenzwert besitzt:

$$\beta_n^V(\delta) \longrightarrow P \left\{ N(0,1) < -z_{1-\alpha} + \frac{\delta}{\sqrt{2\bar{p}}} \right\} \text{ für } n \rightarrow \infty (\delta, \bar{p}, \alpha \text{ fest}).$$

#### 4. Asymptotische relative Effizienz

Population:

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

$$d = p_{12} - p_{21}$$

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

$$H_0: p_{12} \leq p_{21} \quad \text{d.h. } d \leq 0 \quad \text{d.h. } \delta \leq 0$$

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

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

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

Es sei

$n = n_U =$  Stichprobenumfang bei (U)  
(unabhängige Befragungen, Vier-Felder-Test von Fisher);

$n_V =$  Stichprobenumfang bei (V)  
(verbundene Befragungen, Test von McNemar);

$$\lambda_n = \frac{n_V}{n_U} = \frac{n_V}{n}.$$

Wir betrachten nun den Grenzübergang

(5)  $n \rightarrow \infty, \lambda_n \rightarrow \lambda$  ( $p_{11}, p_{22}, \delta$  und  $\alpha$  fest).

Dann gilt:

$$\bar{p} = \frac{1}{2}(p_{12} + p_{21}) = \frac{1}{2}(1 - p_{11} - p_{22}) \quad \text{bleibt fest beim Grenzübergang (5);}$$

$$\bar{\vartheta} = p_{11} + \frac{1}{2}(p_{12} + p_{21}) = p_{11} + \bar{p} \quad \text{bleibt fest beim Grenzübergang (5).}$$

Nun wählen wir beim Grenzübergang (5) den Wert  $\lambda$  so, daß die beiden asymptotischen Gütefunktionen übereinstimmen und beachten, daß beim Grenzprozess zu (V) der Parameter  $\delta$  zu ersetzen ist durch

$$\sqrt{\lambda_n} \cdot \delta = \sqrt{n_V/n} \cdot \sqrt{n} (p_{12} - p_{21}) = \sqrt{n_V} (p_{12} - p_{21}).$$

Die beiden asymptotischen Gütefunktionen sind dann identisch, falls

$$\frac{\sqrt{\lambda}}{\sqrt{2\bar{p}}} = \frac{1}{\sqrt{2\bar{\vartheta}(1-\bar{\vartheta})}}, \quad \text{d.h. falls } \lambda = \frac{\bar{p}}{\bar{\vartheta}(1-\bar{\vartheta})}$$

Somit gilt

$$\text{are}(U : V) = \lim \frac{n_V}{n_U} = \lim \lambda_n = \lambda = \frac{\bar{p}}{\bar{\vartheta}(1-\bar{\vartheta})}$$

Es gilt:  $0 < \text{are}(U : V) \leq 2$ .

*Beispiele:*

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

$$\left. \begin{array}{l} \text{Hier ist: } \bar{p} = \frac{1}{2}(p_{12} + p_{21}) = \frac{1}{4} \\ \bar{\vartheta} = p_{11} + \bar{p} = \frac{1}{2} \end{array} \right\} \Rightarrow \text{are}(U : V) = \frac{1/4}{1/4} = 1$$

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

$$\left. \begin{array}{l} \text{Hier ist: } \bar{p} = \frac{1}{2}(p_{12} + p_{21}) = 0.05 \\ \bar{\vartheta} = p_{11} + \bar{p} = \frac{1}{2} \end{array} \right\} \Rightarrow \text{are}(U : V) = \frac{0.05}{0.25} = \frac{1}{5}$$

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

$$\left. \begin{array}{l} \text{Hier ist: } \bar{p} = \frac{1}{2}(p_{12} + p_{21}) = \frac{1}{2} \\ \bar{\vartheta} = p_{11} + \bar{p} = \frac{1}{2} \end{array} \right\} \Rightarrow \text{are}(U : V) = \frac{1/2}{1/4} = 2$$

## 13. Zur Optimalität von erwartungstreuen Schätzfunktionen

### 1. Ungleichung von Cramér-Rao

$$X \sim F_\vartheta, \vartheta \in \Theta \subset \mathbb{R}$$

$\vartheta$  1-dimensionaler Parameter

$X = (X_1, \dots, X_n)$   $n$ -dimensionale Zufallsgröße

$\mathfrak{X} \subset \mathbb{R}^n$  Stichprobenraum

$T: \mathfrak{X} \rightarrow \mathbb{R}$  Schätzstatistik für  $\vartheta$

Annahme:  $E_\vartheta(T(X)) = \vartheta \quad \forall \vartheta \in \Theta$

d.h.  $T$  ist erwartungstreu für  $\vartheta$

Gütemaß:  $\text{Var}_\vartheta(T(X))$ ;

kleine Varianz: gute Schätzfunktion.

Üblicher Optimalitätsbegriff:

$T$  ist optimal (*effizient*), falls gilt:

- 1)  $T$  ist erwartungstreu für  $\vartheta$ ;
- 2)  $T$  hat kleinstmögliche Varianz unter allen erwartungstreuen Schätzfunktionen.

Log-Likelihood-Funktion:

$$\ell(\vartheta; x) = \begin{cases} \ln p_\vartheta(x) & \text{im diskreten Fall} \\ \ln f_\vartheta(x) & \text{im stetigen Fall} \end{cases}$$

Score-Funktion:

$$\dot{\ell}(\vartheta; x) = \frac{\partial}{\partial \vartheta} \ell(\vartheta; x) = \begin{cases} \frac{\partial}{\partial \vartheta} \ln p_\vartheta(x) = \frac{\dot{p}_\vartheta(x)}{p_\vartheta(x)} & \text{im diskreten Fall} \\ \frac{\partial}{\partial \vartheta} \ln f_\vartheta(x) = \frac{\dot{f}_\vartheta(x)}{f_\vartheta(x)} & \text{im stetigen Fall} \end{cases}$$

Regulärer Fall:

1. Wertebereich der Zufallsgröße  $X$  unabhängig vom Parameter  $\vartheta$ ;
2.  $\Theta =$  offenes Intervall;
3. Dichte  $f_\vartheta(x)$  bzw. Wahrscheinlichkeitsfunktion  $p_\vartheta(x)$  zweimal stetig differenzierbar nach  $\vartheta$ ;
4. Differentiation nach  $\vartheta$  und Integration bzw. Summation bzgl.  $x$  vertauschbar.

Im regulären Fall gilt:

$$E_\vartheta(\dot{\ell}(\vartheta; X)) = 0.$$

Die Fisher-Information ist definiert als

$$J(\vartheta) = E_\vartheta[\dot{\ell}(\vartheta; X)]^2 = \text{Var}_\vartheta(\dot{\ell}(\vartheta; X)) = -E_\vartheta(\ddot{\ell}(\vartheta; X)).$$

Cramér-Rao-Ungleichung:

Wenn  $T: \mathfrak{X} \rightarrow \mathbb{R}$  erwartungstreu ist für  $\vartheta$ , dann gilt  $\text{Var}_\vartheta(T(X)) \geq \frac{1}{J(\vartheta)}$ .

## 2. Satz von Rao

*Nullschätzer:*

Eine Statistik  $T: \mathcal{X} \rightarrow \mathbb{R}$  heißt *Nullschätzer*, falls  $E_{\vartheta}(T(X)) = 0 \quad \forall \vartheta \in \Theta$ .

*Satz von Rao:*

Wenn  $T: \mathcal{X} \rightarrow \mathbb{R}$  erwartungstreu ist für  $\vartheta$ , dann ist  $T$  genau dann effizient, wenn für jeden Nullschätzer  $T_0$  gilt:  $\text{Cov}_{\vartheta}(T(X), T_0(X)) = 0 \quad \forall \vartheta \in \Theta$ .

## 3. Satz von Rao-Blackwell

Es sei

$T: \mathcal{X} \rightarrow \mathbb{R}$  eine suffiziente Statistik für  $\vartheta$ ;

$\hat{\vartheta}: \mathcal{X} \rightarrow \mathbb{R}$  eine erwartungstreue Schätzfunktion für  $\vartheta$ ;

$$\hat{\vartheta}(t) = E_{\vartheta}(\hat{\vartheta}(X) | T(X) = t).$$

Dann ist  $\hat{\vartheta}(t)$  unabhängig von  $\vartheta$ , da  $T$  suffizient ist für  $\vartheta$ , und es gilt:

$$E_{\vartheta}(\hat{\vartheta}(T(X))) = E_{\vartheta}(\hat{\vartheta}(X)) = \vartheta \quad \forall \vartheta \in \Theta$$

$$\text{Var}_{\vartheta}(\hat{\vartheta}(T(X))) \leq \text{Var}_{\vartheta}(\hat{\vartheta}(X)) \quad \forall \vartheta \in \Theta,$$

d.h.  $\hat{\vartheta}(T(X))$  ist eine erwartungstreue Schätzstatistik für  $\vartheta$  mit einer Varianz, die nicht größer sein kann als die Varianz von  $\hat{\vartheta}(X)$ . Bei Vorliegen einer suffizienten Statistik  $T$  ist eine effiziente Schätzfunktion somit faktorisiert über  $T$ .

## 4. Satz von Lehmann-Scheffé

*Vollständige Verteilungsfamilie:*

Es sei  $\mathfrak{F} = \{F_{\vartheta} | \vartheta \in \Theta\}$ . Die Familie  $\mathfrak{F}$  heißt *vollständig*, wenn für jede messbare Funktion

$g: \mathcal{X} \rightarrow \mathbb{R}$  gilt:

$$E_{\vartheta}(g(X)) = 0 \quad \forall \vartheta \in \Theta \quad \Rightarrow \quad g(x) = 0 \quad [\text{f.ü.}].$$

*Vollständige Statistik:*

Es sei

$T: \mathcal{X} \rightarrow \mathbb{R}$  eine Schätzstatistik für  $\vartheta$ ;

$G_{\vartheta}$  die Verteilungsfunktion von  $T(X)$  falls  $X \sim F_{\vartheta}$ ;

$$\mathfrak{F}_T = \{G_{\vartheta} | \vartheta \in \Theta\}.$$

Die Statistik  $T$  heißt *vollständig*, wenn die Verteilungsfamilie  $\mathfrak{F}_T$  vollständig ist.

*Satz von Lehmann-Scheffé:*

Es sei  $T: \mathcal{X} \rightarrow \mathbb{R}$  eine vollständige suffiziente Statistik für  $\vartheta$ ;

$\hat{\vartheta}: \mathcal{X} \rightarrow \mathbb{R}$  eine erwartungstreue Schätzfunktion für  $\vartheta$ ;

$$\hat{\vartheta}(t) = E_{\vartheta}(\hat{\vartheta}(X) | T(X) = t).$$

Dann ist  $\hat{\vartheta}(T(X))$  effizient für  $\vartheta$  (erwartungstreu mit kleinstmöglicher Varianz), und im wesentlichen ist dies die einzige effiziente Schätzung für  $\vartheta$ .

## 5. Zum durchschnittlichen quadratischen Fehler

Es sei

$T: \mathcal{X} \rightarrow \mathbb{R}$  eine Schätzstatistik für  $\vartheta$  (nicht notwendigerweise erwartungstreu).

Dann ist der *durchschnittliche quadratische Fehler* (MSE = mean square error) definiert als

$$\text{MSE}_{\vartheta}(T(X)) = \mathbb{E}_{\vartheta}(T(X) - \vartheta)^2.$$

Falls  $T$  erwartungstreu ist für  $\vartheta$ , so gilt

$$\mathbb{E}_{\vartheta}(T(X)) = \vartheta$$

$$\text{Var}_{\vartheta}(T(X)) = \mathbb{E}_{\vartheta}(T(X) - \vartheta)^2 = \text{MSE}_{\vartheta}(T(X)).$$

Falls  $T$  nicht erwartungstreu ist für  $\vartheta$ , so gilt

$$\mathbb{E}_{\vartheta}(T(X)) = m(\vartheta);$$

$$m(\vartheta) - \vartheta = b(\vartheta) = \text{Verzerrung (= Bias) von } T;$$

$$\begin{aligned} \text{MSE}_{\vartheta}(T(X)) &= \mathbb{E}_{\vartheta}(T(X) - \vartheta)^2 \\ &= \mathbb{E}_{\vartheta}(T(X) - m(\vartheta))^2 + (m(\vartheta) - \vartheta)^2 \\ &= \text{Var}_{\vartheta}(T(X)) + b^2(\vartheta). \end{aligned}$$

Es gibt Schätzfunktionen, bei denen der durchschnittliche quadratische Fehler MSE kleiner ist als die Varianz der effizienten *erwartungstreuen* Schätzfunktion!

## 14. Informationsmatrix bei mehrdimensionalem Parameter

### 1. Ausgangssituation

$$X \sim F_{\vartheta}, \vartheta \in \Theta \left( \subset \mathbb{R}^k \right);$$

$\vartheta = (\vartheta_1, \dots, \vartheta_k)$   $k$ -dimensionaler Parameter;

$\Theta \left( \subset \mathbb{R}^k \right)$  offenes  $k$ -dimensionales Intervall;

$X = (X_1, \dots, X_n)$   $n$ -dimensionale Zufallsgröße;

$\mathfrak{X} \left( \subset \mathbb{R}^n \right)$  Stichprobenraum;

$T: \mathfrak{X} \rightarrow \mathbb{R}^k$  Schätzstatistik für  $\vartheta$ .

### 2. Mathematische Grundlagen

Lit.: Witting (1985). Mathematische Statistik I; S.153 ff.

$\gamma: \Theta \rightarrow \mathbb{R}$  d.h.  $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \Theta \mapsto \gamma(\vartheta) \in \mathbb{R}$ .

*Gradient von  $\gamma$ :*

$\frac{\partial \gamma}{\partial \vartheta_i}$  = partielle Ableitung von  $\gamma$  nach  $\vartheta_i$ ;

$$\frac{\partial \gamma}{\partial \vartheta} = \left( \frac{\partial \gamma}{\partial \vartheta_1}, \dots, \frac{\partial \gamma}{\partial \vartheta_k} \right)^T = (k \times 1) = \text{Gradient von } \gamma.$$

*Hesse-Matrix:*

$\frac{\partial^2 \gamma}{\partial \vartheta_i \partial \vartheta_j}$  = 2-fache partielle Ableitung von  $\gamma$  nach  $\vartheta_i$  und  $\vartheta_j$ ;

$$\frac{\partial^2 \gamma}{\partial \vartheta^2} = \left( \frac{\partial^2 \gamma}{\partial \vartheta_i \partial \vartheta_j} \right) = (k \times k) = \text{Hesse-Matrix (symmetrisch)}.$$

*Jacobi-Matrix (Funktionalmatrix):*

$$\gamma: \Theta \rightarrow \mathbb{R}^h;$$

$\gamma = (\gamma_1, \dots, \gamma_h)^T$  = Vektor mit  $h$  Komponenten;

$$\gamma_i(\vartheta) = \gamma_i(\vartheta_1, \dots, \vartheta_k);$$

$\frac{\partial \gamma_i}{\partial \vartheta_j}$  = partielle Ableitung von  $\gamma_i$  nach  $\vartheta_j$  ( $i = 1, \dots, h; j = 1, \dots, k$ );

$$\frac{\partial \gamma}{\partial \vartheta} = \left( \frac{\partial \gamma_i}{\partial \vartheta_j} \right) = (h \times k) = \text{Jacobi-Matrix (Funktionalmatrix)}.$$

### 3. (Fisher-) Informationsmatrix

$$X \sim F_\vartheta, \vartheta \in \Theta \left( \subset \mathbb{R}^k \right);$$

$\mathfrak{F} = \{F_\vartheta \mid \vartheta \in \Theta\}$  = Familie aller zulässigen Verteilungen;

stetiger Fall:  $f_\vartheta$  = Dichte zu  $F_\vartheta$ ;

diskreter Fall:  $p_\vartheta$  = Wahrscheinlichkeitsfunktion zu  $F_\vartheta$ ;

Regulärer Fall:

1. Wertebereich der Zufallsgröße  $X$  unabhängig vom Parameter  $\vartheta$ ;
2.  $\Theta$  = offenes  $k$ -dimensionales Intervall;
3. Dichte  $f_\vartheta(x)$  bzw. Wfunktion  $p_\vartheta(x)$  zweimal stetig differenzierbar nach  $\vartheta$ ;
4. Differentiation nach  $\vartheta$  und Integration bzw. Summation bzgl.  $x$  vertauschbar.

Log-Likelihood-Funktion:

$$\ell(\vartheta; x) = \begin{cases} \ln p_\vartheta(x) & \text{im diskreten Fall} \\ \ln f_\vartheta(x) & \text{im stetigen Fall} \end{cases}$$

Score-Vektor:

$$\dot{\ell}(\vartheta; x) = \frac{\partial}{\partial \vartheta} \ell(\vartheta; x) = (k \times 1).$$

Im regulären Fall gilt:

$$E_\vartheta \left( \dot{\ell}(\vartheta; X) \right) = 0.$$

Es sei  $\dot{\ell} = \dot{\ell}(\vartheta; X) = (k \times 1)$ . Dann ist die (Fisher-) Informationsmatrix definiert als

$$\begin{aligned} \mathbf{J}(\vartheta) &= E_\vartheta \left[ \dot{\ell} \dot{\ell}^\top \right] = \text{Cov}_\vartheta \left( \dot{\ell} \right) = (k \times k) \\ &= \text{Kovarianz-Matrix des Score-Vektors } \dot{\ell} = \dot{\ell}(\vartheta; X). \end{aligned}$$

Im regulären Fall kann die Informationsmatrix auch geschrieben werden als:

$$\mathbf{J}(\vartheta) = -E_\vartheta \left( \frac{\partial^2}{\partial \vartheta^2} \ell(\vartheta; X) \right) = -E_\vartheta \left( \ddot{\ell}(\vartheta; X) \right).$$

### 4. Cramér-Rao-Ungleichung im mehrdimensionalen Fall

a) Vergleich von Kovarianz-Matrizen (Löwner-Ordnung):

Lit.: Witting (1985). Mathematische Statistik I, S. 59.

$\hat{\vartheta}$ :  $\mathfrak{X} \rightarrow \mathbb{R}^k$  erwartungstreue Schätzstatistik für  $\vartheta$ ;

$\hat{\hat{\vartheta}}$ :  $\mathfrak{X} \rightarrow \mathbb{R}^k$  andere erwartungstreue Schätzstatistik für  $\vartheta$ ;

$$\mathbf{C}_1 = \text{Cov}_\vartheta \left( \hat{\vartheta}(X) \right) = E_\vartheta \left( \hat{\vartheta} - \vartheta \right) \left( \hat{\vartheta} - \vartheta \right)^\top = (k \times k);$$

$$\mathbf{C}_2 = \text{Cov}_\vartheta \left( \hat{\hat{\vartheta}}(X) \right) = E_\vartheta \left( \hat{\hat{\vartheta}} - \vartheta \right) \left( \hat{\hat{\vartheta}} - \vartheta \right)^\top = (k \times k).$$

$$\hat{\vartheta} \text{ besser als } \hat{\hat{\vartheta}} \Leftrightarrow \text{Var} \left( u^\top \hat{\vartheta} \right) \leq \text{Var} \left( u^\top \hat{\hat{\vartheta}} \right) \quad \forall u = (k \times 1)$$

$$\Leftrightarrow \mathbf{C}_2 - \mathbf{C}_1 \text{ positiv semidefinit}$$

$$\Leftrightarrow \mathbf{C}_1 \leq \mathbf{C}_2 \text{ (Definition der Löwner-Ordnung).}$$

b) Mehrdimensionale Cramér-Rao-Ungleichung:

Lit.: Witting (1985). Mathematische Statistik I, S. 317 ff.

Es sei

$\hat{\vartheta}: \mathcal{X} \rightarrow \mathbb{R}^k$  eine erwartungstreue Schätzstatistik für  $\vartheta$ ;

$\mathbf{C} = \text{Cov}_{\vartheta}(\hat{\vartheta}(X)) = E_{\vartheta}(\hat{\vartheta} - \vartheta)(\hat{\vartheta} - \vartheta)^{\top} = \text{Kovarianzmatrix von } \hat{\vartheta}$ ;

$\mathbf{J}(\vartheta)$  die Informationsmatrix (wie oben).

Dann gilt:

$\mathbf{C} \geq \mathbf{J}^{-1}(\vartheta)$  (im Sinne der Löwner-Ordnung).

Die inverse Informationsmatrix ist eine untere Schranke (im Sinne der Löwner-Ordnung) für die Kovarianzmatrix eines beliebigen erwartungstreuen Schätzvektors  $\hat{\vartheta}$ .

## 5. Asymptotische Verteilung von Maximum-Likelihood-Schätzungen

Lit.: Cramér (1945). Mathematical Methods of Statistics, S. 500 ff.

Es sei

$X_1, \dots, X_n$  unabhängig, je mit der gleichen Verteilungsfunktion  $F_{\vartheta}$ ,  $\vartheta \in \Theta \left( \subset \mathbb{R}^k \right)$ ;

stetiger Fall:  $f_{\vartheta} = \text{Dichte zu } F_{\vartheta}$ ;

diskreter Fall:  $p_{\vartheta} = \text{Wahrscheinlichkeitsfunktion zu } F_{\vartheta}$ ;

$$\ell(\vartheta; x) = \begin{cases} \ln p_{\vartheta}(x) & \text{im diskreten Fall} \\ \ln f_{\vartheta}(x) & \text{im stetigen Fall} \end{cases}$$

= Log-Likelihoodfunktion der Komponente  $X_i$

$\mathbf{J}(\vartheta)$  die zugehörige Informationsmatrix;

$\hat{\vartheta}$  die Maximum-Likelihood-Schätzung für  $\vartheta$ .

Dann gilt im regulären Fall:

$$\sqrt{n}(\hat{\vartheta} - \vartheta) \xrightarrow[n \rightarrow \infty]{d} N_k(\mathbf{0}, \Sigma) \quad \text{wobei } \Sigma = \mathbf{J}^{-1}(\vartheta).$$

Eine Maximum-Likelihood-Schätzung ist also im regulären Fall

- konsistent,
- asymptotisch erwartungstreu,
- asymptotisch normalverteilt und
- asymptotisch effizient (asymptotische Kovarianzmatrix = inverse Informationsmatrix = Cramér-Rao-Schranke).

## 15. Zur empirischen Verteilungsfunktion

### 1. Einführung

$X$  (eindimensionale) Zufallsgröße mit beliebiger Verteilungsfunktion  $F$ .

Ziel: Verteilungsfunktion  $F$  schätzen.

*Analoges Problem* für Wahrscheinlichkeiten:

A beliebiges zufälliges Ereignis mit unbekannter Wahrscheinlichkeit  $P(A)$ .

Ziel:  $P(A)$  schätzen.

Klassisches Vorgehen:

Zugrunde liegendes Experiment  $n$ -mal durchführen;

$Y$  = Anzahl der Versuche, bei denen  $A$  eintritt;

dann gilt:  $Y \sim \text{Bi}(n, \vartheta)$  mit  $\vartheta = P(A)$ ;

$$Y/n \xrightarrow[n \rightarrow \infty]{P} \vartheta;$$

Vertrauensgrenzen für  $\vartheta$  mit bekannten Verfahren bei Binomialverteilung.

Nun wollen wir analog vorgehen, um die Verteilungsfunktion  $F$  zu schätzen.

### 2. Empirische Verteilungsfunktion

$X$  (eindimensionale) Zufallsgröße mit beliebiger Verteilungsfunktion  $F$ .

Zugrunde liegendes Experiment  $n$ -mal durchführen;

→  $X_1, \dots, X_n$  unabhängig, je mit Verteilungsfunktion  $F$ ;

$x_1, \dots, x_n$  zugehörige Realisationen.

*Empirische Verteilungsfunktion zur Stichprobe*  $x_1, \dots, x_n$ :

$$\begin{aligned} F_n(x) &= \frac{1}{n} \cdot \#\{i \mid x_i \leq x\} \quad \text{mit } x \in \mathbb{R} \\ &= \frac{1}{n} \cdot \text{Anzahl der } x_i \text{ mit } x_i \leq x. \end{aligned}$$

*Empirische Verteilungsfunktion zu den Zufallsgrößen*  $X_1, \dots, X_n$ :

$$F_n(x) = \frac{1}{n} \cdot \#\{i \mid X_i \leq x\} \quad \text{mit } x \in \mathbb{R}.$$

(= empirische Verteilungsfunktion als zufällige Funktion!)

### 3. Satz von Glivenko-Cantelli (Hauptsatz der Statistik)

Es sei  $F_n$  die empirische Verteilungsfunktion zu  $X_1, \dots, X_n$ . Dann gilt aufgrund des schwachen Gesetzes der großen Zahlen;

$$F_n(x) \xrightarrow[n \rightarrow \infty]{P} F(x) \quad \text{für jedes } x \in \mathbb{R}, \text{ d.h.}$$

$$\left| F_n(x) - F(x) \right| \xrightarrow[n \rightarrow \infty]{P} 0 \quad \text{für jedes } x \in \mathbb{R}.$$

*Satz von Glivenko-Cantelli:*

Für jede beliebige (stetige oder diskrete) Verteilungsfunktion  $F$  gilt:

$$\sup_{x \in \mathbb{R}} \left| F_n(x) - F(x) \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Anwendung:

a) Beispiel:

$$\mu = \int x dF(x) = \text{Erwartungswert von } X;$$

$$\hat{\mu} = \int x dF_n(x) = \frac{1}{n} (X_1 + \dots + X_n) = \bar{X}.$$

Wenn in der Definition des Parameters (Funktionals)  $\mu = \mu(F)$  die unbekannte Verteilungsfunktion  $F$  ersetzt wird durch die empirische Verteilungsfunktion  $F_n$ , dann erhält man als Schätzfunktion  $\hat{\mu} = \mu(F_n) = \bar{X}$ .

b) Eine allgemeine Schätzmethode (sogenannte *Plug-in Methode*):

Es sei  $\vartheta = \vartheta(F)$  ein Funktional von  $F$  (beliebiger Parameter von  $F$ ). Wenn wir nun in der Definition des Funktionals  $\vartheta = \vartheta(F)$  die unbekannte Verteilungsfunktion  $F$  ersetzen durch die empirische Verteilungsfunktion  $F_n$ , dann erhalten wir für den unbekannt Parameter  $\vartheta = \vartheta(F)$  die Schätzfunktion  $\hat{\vartheta} = \vartheta(F_n)$ . Da  $F_n \rightarrow F$  für  $n \rightarrow \infty$  aufgrund des Satzes von Glivenko-Cantelli, so kann man erwarten, daß auch  $\hat{\vartheta} = \vartheta(F_n) \rightarrow \vartheta(F) = \vartheta$  für  $n \rightarrow \infty$ ; die Schätzung  $\hat{\vartheta} = \vartheta(F_n)$  wird also im allgemeinen zumindest konsistent sein.

#### 4. Satz von Kolmogoroff

Es seien

$X_1, \dots, X_n$  unabhängige Zufallsgrößen je mit gleicher Verteilungsfunktion  $F$ ;

$F_n$  die zugehörige empirische Verteilungsfunktion;

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|;$$

$$D_n^+ = \sup_{x \in \mathbb{R}} (F_n(x) - F(x));$$

$$D_n^- = \sup_{x \in \mathbb{R}} (F(x) - F_n(x)).$$

Falls  $F$  stetig ist, so ist die Verteilung dieser drei Statistiken unabhängig von der speziellen Form von  $F$ , und asymptotisch gilt:

$$\sqrt{n} D_n \xrightarrow[n \rightarrow \infty]{d} K(\cdot);$$

$$\sqrt{n} D_n^+ \xrightarrow[n \rightarrow \infty]{d} L(\cdot);$$

$$\sqrt{n} D_n^- \xrightarrow[n \rightarrow \infty]{d} L(\cdot);$$

dabei sind die beiden Grenzverteilungen  $K$  (= Kolmogoroff-Verteilung) und  $L$  definiert als

$$K(y) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 y^2} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 y^2} \quad (\text{für } y > 0);$$

$$L(y) = 1 - e^{-2y^2} \quad (\text{für } y > 0).$$

*Bemerkung zum Beweis:*

1° Wenn  $X$  eine Zufallsgröße ist mit der stetigen Verteilungsfunktion  $F$ , dann besitzt die Zufallsgröße  $Y = F(X)$  eine Gleichverteilung (Rechteckverteilung) im Intervall  $(0,1)$ :  $Y \sim U(0,1)$ .

2° Es seien

$X_1, \dots, X_n$  unabhängige Zufallsgrößen je mit der gleichen stetigen Verteilungsfunktion  $F$ ;  
 $Y_i = F(X_i)$ , für  $i = 1, \dots, n$ .

Dann sind  $Y_1, \dots, Y_n$  unabhängige Zufallsgrößen je mit einer Gleichverteilung  $U(0,1)$ . Es seien  $F_n$  und  $G_n$  die empirischen Verteilungsfunktionen zu  $X_1, \dots, X_n$  bzw.  $Y_1, \dots, Y_n$ , und  $G$  sei die Verteilungsfunktion zu  $U(0,1)$ . Dann gilt  $F_n(x) - F(x) = G_n(y) - G(y)$ , wobei  $y = F(x)$ . Somit bleiben die Abstände  $D_n, D_n^+$  und  $D_n^-$  invariant bei der Transformation  $Y_i = F(X_i)$ ,  $i = 1, \dots, n$ .

*Zur praktischen Durchführung:*

Die exakte Verteilung der Statistiken  $D_n, D_n^+$  und  $D_n^-$  für endliches  $n$  im stetigen Fall findet man in StatXact (Statistics, One sample, Kolmogorov ...). Man beachte, dass eine Zufallsgröße  $X$  mit einer beliebigen stetigen Verteilungsfunktion  $F$  durch  $Y = F(X)$  in eine Zufallsgröße  $Y$  mit einer Gleichverteilung  $U(0,1)$  transformiert werden kann. Falls  $F$  Sprünge aufweist, so hängt die Verteilung der Statistiken  $D_n, D_n^+$  und  $D_n^-$  von der speziellen Form von  $F$  ab. Die exakte Verteilung für einige ausgewählte diskrete Verteilungen findet man wiederum in StatXact (Statistics, One sample, Kolmogorov, ...).

## 5. Test von Kolmogoroff (Einstichprobenproblem)

$X_1, \dots, X_n$  unabhängige Zufallsgrößen je mit gleicher Verteilungsfunktion  $F$ ;

$H_0$ :  $F \equiv F_0$  ( $F_0 =$  vorgegebene *stetige* Verteilungsfunktion);

$H_1$ : a)  $F \not\equiv F_0$

b)  $F(x) \geq F_0(x) \forall x \in \mathbb{R}$  (und  $F \not\equiv F_0$ )

c)  $F(x) \leq F_0(x) \forall x \in \mathbb{R}$  (und  $F \not\equiv F_0$ )

Berechne die Teststatistiken

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|;$$

$$D_n^+ = \sup_{x \in \mathbb{R}} (F_n(x) - F_0(x));$$

$$D_n^- = \sup_{x \in \mathbb{R}} (F_0(x) - F_n(x)).$$

Unter  $H_0$  ist die Verteilung dieser drei Statistiken unabhängig von der speziellen Form von  $F_0$ .

Testvorschrift im Fall a):

Berechne für die vorliegende Stichprobe  $x_1, \dots, x_n$  den Wert der Teststatistik  $D_n$ :

$$d_{\text{obs}} = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \quad \text{wobei } F_n = \text{empirische Verteilungsfkt. zu } x_1, \dots, x_n;$$

berechne dann den exakten  $p$ -Wert

$$p_0 = P_0 \{D_n \geq d_{\text{obs}}\};$$

verwirf  $H_0$ , falls  $p_0 \leq \alpha$ .

Analog für die Probleme b) und c).

Praktische Durchführung mit StatXact (Statistics, One sample, Kolmogorov ...). Auch im diskreten Fall (d.h. falls  $F_0$  diskret) berechnet StatXact die exakte Lösung für einige ausgewählte Verteilungen.

## 6. Satz von Kolmogoroff-Smirnoff

Es seien

$$\left. \begin{array}{l} X_1, \dots, X_{n_1} \\ Y_1, \dots, Y_{n_2} \end{array} \right\} \text{unabhängig, je mit gleicher Verteilungsfunktion } F;$$

$F_{n_1}$  die empirische Verteilungsfunktion zu  $X_1, \dots, X_{n_1}$ ;

$G_{n_2}$  die empirische Verteilungsfunktion zu  $Y_1, \dots, Y_{n_2}$ ;

$$D = \sup_{x \in \mathbb{R}} |F_{n_1}(x) - G_{n_2}(x)|;$$

$$D^+ = \sup_{x \in \mathbb{R}} (F_{n_1}(x) - G_{n_2}(x));$$

$$D^- = \sup_{x \in \mathbb{R}} (G_{n_2}(x) - F_{n_1}(x)).$$

Falls  $F$  stetig ist, so ist die Verteilung dieser drei Statistiken unabhängig von der speziellen Form von  $F$ ; sie hängt nur ab von den beiden Stichprobenumfängen  $n_1$  und  $n_2$ . Asymptotisch gilt:

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} D \xrightarrow[n_1, n_2 \rightarrow \infty]{d} K(\cdot);$$

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} D^+ \xrightarrow[n_1, n_2 \rightarrow \infty]{d} L(\cdot);$$

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} D^- \xrightarrow[n_1, n_2 \rightarrow \infty]{d} L(\cdot);$$

dabei sind die beiden Grenzverteilungen  $K$  und  $L$  definiert wie oben beim Satz von Kolmogoroff.

Die exakte Verteilung der Statistiken  $D, D^+$  und  $D^-$  für endliche Stichprobenumfänge  $n_1$  und  $n_2$  kann man berechnen lassen mit StatXact (Statistics, Two independent samples, Kolmogorov-Smirnov, ...). Falls  $F$  Sprünge aufweist, so hängt die Verteilung der Statistiken  $D, D^+$  und  $D^-$  von der speziellen Form von  $F$  ab, und es können dann Bindungen auftreten. In diesem Fall berechnet StatXact die exakte bedingte Verteilung bei gegebener Bindungsstruktur.

## 7. Test von Kolmogoroff-Smirnoff (Zweistichprobenproblem)

$$\left. \begin{array}{l} X_1, \dots, X_{n_1} \\ Y_1, \dots, Y_{n_2} \end{array} \right\} \text{unabhängig, } \begin{cases} \text{je mit der gleichen Verteilungsfunktion } F \\ \text{je mit der gleichen Verteilungsfunktion } G \end{cases}$$

$H_0: F \equiv G$  (= beliebige Verteilungsfunktion)

$H_1: a) F \not\equiv G$

b)  $F(x) \geq G(x) \quad \forall x \in \mathbb{R}$  (und  $F \not\equiv G$ )

c)  $F(x) \leq G(x) \quad \forall x \in \mathbb{R}$  (und  $F \not\equiv G$ )

Unter  $H_0$  ist die exakte Verteilung der drei Statistiken  $D$ ,  $D^+$  und  $D^-$  berechenbar (bei Bindungen: exakte bedingte Verteilung bei gegebener Bindungsstruktur).

Testvorschrift im Fall a):

$F_{n_1}$  die empirische Verteilungsfunktion zur vorliegenden Stichprobe  $x_1, \dots, x_{n_1}$ ;

$G_{n_2}$  die empirische Verteilungsfunktion zur vorliegenden Stichprobe  $y_1, \dots, y_{n_2}$ ;

Berechne den Wert der Teststatistik  $D$ :

$$d_{\text{obs}} = \sup_{x \in \mathbb{R}} |F_{n_1}(x) - G_{n_2}(x)|;$$

berechne dann den exakten  $p$ -Wert

$$p_0 = P_0\{D \geq d_{\text{obs}}\};$$

verwirf  $H_0$ , falls  $p_0 \leq \alpha$ .

Analog für die Probleme b) und c).

Praktische Durchführung mit StatXact:

*Statistics, Two independent samples, Kolmogorov-Smirnov, ...*

## 16. Über Dichteschätzung

### 1. Einführung

$X_1, \dots, X_n$  unabhängig, je mit gleicher Verteilungsfunktion  $F$ ;

$F$  unbekannt;

$f$  Dichte von  $F$ ,  $f$  stetig;

$F$  schätzen:  $F_n =$  empirische Verteilungsfunktion;

$$F_n(x) \xrightarrow[n \rightarrow \infty]{P} F(x);$$

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow[n \rightarrow \infty]{d} N(0, \tau^2) \text{ mit } \tau^2 = F(x)(1 - F(x)).$$

$f$  schätzen: ???

### 2. Histogramm-Schätzer

Annahme: Wertebereich von  $X_i: [A, B]$  mit  $-\infty < A < B < \infty$ ;

Wertebereich der  $X_i$  zerlegen in  $k$  Intervalle gleicher Länge:  $I_1, I_2, \dots, I_k$ ;

Intervall-Länge:  $b = (B - A)/k$ ;

$\vartheta_j = P\{X \in I_j\} = \int_{I_j} f(x) dx \approx b \cdot f_j$ , wobei  $f_j =$  Dichte in der Intervallmitte von  $I_j$ ;

$n_j = \#\{i \mid X_i \in I_j\} =$  absolute Häufigkeit im Intervall  $I_j$  ( $i = 1, \dots, n; j = 1, \dots, k$ );

$n_j \sim \text{Bi}(n, \vartheta_j)$ ;

$$\frac{n_j}{n} \xrightarrow[n \rightarrow \infty]{P} \vartheta_j \approx b \cdot f_j;$$

$\hat{f}_j = \frac{n_j}{b \cdot n}$  Histogrammschätzung für  $f_j$  (= Dichte in Intervallmitte von  $I_j$ ).

Falls

$$\begin{aligned} n &\rightarrow \infty \\ k &\rightarrow \infty \text{ (d.h. } b \rightarrow 0) \\ n_j &\rightarrow \infty \text{ (d.h. } n\vartheta_j \rightarrow \infty \text{ bzw. } nb \rightarrow \infty), \end{aligned}$$

so gilt:

$$I_j \rightarrow \{x\};$$

$$\hat{f}_j \xrightarrow{P} f(x);$$

$$\sqrt{n_j}(\hat{f}_j - f(x)) \xrightarrow{d} N(0, \sigma^2(x)).$$

Eine verbreitete Faustregel besagt, daß  $k \approx \sqrt{n}$  betragen sollte. In diesem Fall gilt für die zentralen Intervalle  $n_j \approx n/k \approx \sqrt{n}$ , und die Konvergenzgeschwindigkeit des Dichteschätzers  $\hat{f}_j$  beträgt somit  $\sqrt{n_j} \approx n^{\frac{1}{4}}$ , während die Konvergenzgeschwindigkeit im Falle der empirischen Verteilungsfunktion  $\sqrt{n}$  beträgt.

### 3. Kern-Dichteschätzer

$x_1, \dots, x_n$  Realisationen zu  $X_1, \dots, X_n$ ;

Beispiel einer Kernfunktion:

$$K(x) = \begin{cases} 1 & \text{für } x \in \left[-\frac{1}{2}, +\frac{1}{2}\right] \\ 0 & \text{sonst;} \end{cases}$$

$K =$  Rechteckskern auf  $\left[-\frac{1}{2}, +\frac{1}{2}\right]$ ;

$$K_b(x) = \frac{1}{b} K\left(\frac{x}{b}\right) = \begin{cases} \frac{1}{b} & \text{für } x \in \left[-\frac{b}{2}, +\frac{b}{2}\right] \\ 0 & \text{sonst;} \end{cases}$$

$K_b =$  Kern  $K$  mit Bandbreite  $b$ .

Allgemeine Kernfunktion:

$$K: \mathbb{R} \rightarrow [0, \infty) \quad \text{mit} \quad \int K(x) dx = 1;$$

$$K_b(x) = \frac{1}{b} K\left(\frac{x}{b}\right) \quad \text{Kern } K \text{ mit Bandbreite } b \text{ (} b = \text{Skalenparameter).}$$

Kern-Dichteschätzer:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_b(x - x_i).$$

Es gilt:

$$\int_{-\infty}^{\infty} \hat{f}(x) dx = \frac{1}{n} \sum_{i=1}^n \underbrace{\int_{-\infty}^{\infty} K_b(x - x_i) dx}_{=1} = 1;$$

für  $b \rightarrow 0$ :  $\hat{f} \rightarrow$  Nadelplot (zu zackig, undersmoothing);

für  $b \rightarrow \infty$ :  $\hat{f}$  verflacht zu 0 (zu glatt, oversmoothing).

Falls

$$n \rightarrow \infty, \quad b \rightarrow 0 \quad \text{und} \quad nb \rightarrow \infty,$$

so gilt:

$$\hat{f}(x) \xrightarrow{P} f(x).$$

Das Konvergenzverhalten des Kern-Dichteschätzers und des Histogrammschätzers ist im wesentlichen gleich.

## 17. Invarianzprinzipien, Pitman-Schätzer

*Literatur:*

Witting (1985). Mathematische Statistik I, Seite 436-446.

### 1. Lagemodelle

$X_1, \dots, X_n$  unabhängig, je mit gleicher Dichtefunktion  $f_\vartheta$ ;

$$f_\vartheta(x) = f(x - \vartheta) \quad f \text{ bekannt, } \vartheta \in \mathbb{R}.$$

Man spricht dann von einem *Lagemodell (Lokationsmodell)*, weil nur der Lageparameter  $\vartheta$  unbekannt ist.

Wir wollen den Lageparameter  $\vartheta$  optimal schätzen aufgrund einer Stichprobe  $x_1, \dots, x_n$ . Dabei soll die Schätzung äquivariant bleiben bei Lageänderungen, d.h.

$$\hat{\vartheta}(x_1 + d, \dots, x_n + d) = \hat{\vartheta}(x_1, \dots, x_n) + d \quad \forall d \in \mathbb{R}.$$

*Optimalitätsforderung an  $\hat{\vartheta}$*

- 1)  $\hat{\vartheta}$  soll äquivariant sein bei Lageänderungen;
- 2)  $\text{MSE}(\hat{\vartheta}) = E_\vartheta(\hat{\vartheta} - \vartheta)^2 \stackrel{!}{=} \min.$

### 2. Äquivalenzklassen, Translationsbahnen

Bei einem Lagemodell ist der Stichprobenraum zu  $X_1, \dots, X_n$  stets gegeben durch  $\mathfrak{X} = \mathbb{R}^n$ . Es sei  $t \in \mathbb{R}^n$  und

$$\mathfrak{X}_t = \left\{ (x_1, \dots, x_n) \mid x_i - t_i = c \quad \forall i \right\},$$

d.h.

$$x = (x_1, \dots, x_n) \in \mathfrak{X}_t \Leftrightarrow \exists c \in \mathbb{R} \text{ mit } x_i = t_i + c \quad \forall i.$$

Die Menge  $\mathfrak{X}_t$  heißt daher *Translationsbahn* zu  $t$ . Die Menge  $\mathfrak{X}_t$  nennt man auch eine Äquivalenzklasse. Wenn zwei Punkte  $x$  und  $\tilde{x}$  zur gleichen Äquivalenzklasse (Translationsbahn) gehören, so schreiben wir  $x \equiv \tilde{x}$  (lies:  $x$  ist äquivalent zu  $\tilde{x}$ ). Es gilt also

$$\begin{aligned} x \equiv \tilde{x} &\Leftrightarrow \exists t \in \mathfrak{X} \text{ mit } x, \tilde{x} \in \mathfrak{X}_t \\ &\Leftrightarrow \exists c \in \mathbb{R} \text{ mit } x_i - \tilde{x}_i = c \quad \forall i. \end{aligned}$$

*Äquivalenzklassen-Statistik*

Wir betrachten die Statistik  $T: \mathfrak{X} \rightarrow \mathfrak{X}$ , welche definiert ist durch

$$x = (x_1, \dots, x_n) \rightarrow T(x) = t = (0, x_2 - x_1, x_3 - x_1, \dots, x_n - x_1).$$

Es gilt offensichtlich

$$\begin{aligned} T(x) = T(\tilde{x}) &\Leftrightarrow x_i - x_1 = \tilde{x}_i - \tilde{x}_1 \quad \forall i \\ &\Leftrightarrow x_i - \tilde{x}_i = x_1 - \tilde{x}_1 = c \quad \forall i \\ &\Leftrightarrow x \equiv \tilde{x}. \end{aligned}$$

Die Statistik  $T$  ordnet jeder Äquivalenzklasse einen (eindeutig bestimmten) Repräsentanten  $t$  zu.  $T$  heißt daher eine *Äquivalenzklassen-Statistik*. Wir fragen uns, ob es weitere Äquivalenzklassenstatistiken gibt. Es sei  $\tilde{T}: \mathfrak{X} \rightarrow \mathfrak{X}$  eine beliebige Äquivalenzklassen-Statistik, d.h. für alle  $x \in \mathfrak{X}_t$  gilt  $\tilde{T}(x) = \tilde{t} \in \mathfrak{X}_t$ . Dabei sei  $t$  der Repräsentant der Statistik  $T$ . Da nun beide Repräsentanten  $t$  und  $\tilde{t}$  zur gleichen Äquivalenzklasse gehören, so gilt  $t_i - \tilde{t}_i = c$  für  $i = 1, \dots, n$ . Die beiden Repräsentanten der Statistiken  $T$  und  $\tilde{T}$  können sich also nur durch eine Translation unterscheiden.

#### *Invariante eindimensionale Statistiken*

Eine Statistik  $h: \mathfrak{X} \rightarrow \mathbb{R}$  heißt *invariant*, falls  $h(x) = h(\tilde{x})$  für  $x \equiv \tilde{x}$ . Eine invariante Statistik besitzt also für alle Werte einer Translationsbahn den gleichen Wert, d.h.

für alle  $x \in \mathfrak{X}_t$ . Jede

invariante Statistik kann somit dargestellt werden als eine Funktion der Äquivalenzklassen-Statistik:  

$$h(X) = h(T(X)).$$

#### *Zerlegung einer äquivarianten Schätzfunktion*

Wir suchen nun eine vereinfachende Darstellung einer äquivarianten Schätzfunktion  $\hat{\vartheta}: \mathfrak{X} \rightarrow \mathbb{R}$ .

Einfache äquivariante Statistiken sind beispielsweise:

$$n-1$$

Es sei nun

$\hat{\vartheta}: \mathfrak{X} \rightarrow \mathbb{R}$  eine äquivariante Schätzfunktion;

$e: \mathfrak{X} \rightarrow \mathbb{R}$  eine (einfache) äquivariante Statistik.

Dann ist offensichtlich  $h(x) = \hat{\vartheta}(x) - e(x)$  eine invariante Statistik und daher darstellbar als

$h(x) = h(T(x))$ . Damit kann jede äquivariante Schätzfunktion  $\hat{\vartheta}: \mathfrak{X} \rightarrow \mathbb{R}$  dargestellt werden als

$$\hat{\vartheta}(x) = e(x) + h(x),$$

wobei

$e: \mathfrak{X} \rightarrow \mathbb{R}$  eine (einfache) äquivariante Statistik und

$h: \mathfrak{X} \rightarrow \mathbb{R}$  eine invariante Statistik ist.

### **3. Lösung des Optimalitätsproblems, Teil 1**

Es gilt

$$\text{MSE}(\hat{\vartheta}(X)) = E_{\vartheta}(\hat{\vartheta}(X) - \vartheta)^2 = \int_{\mathfrak{X}} (\hat{\vartheta}(x) - \vartheta)^2 f_{\vartheta}^{(n)}(x) dx,$$

wobei

$$f_{\vartheta}^{(n)}(x) = f_{\vartheta}(x_1) \cdots f_{\vartheta}(x_n) = f(x_1 - \vartheta) \cdots f(x_n - \vartheta) = f^{(n)}(x - \vartheta_v)$$

$$x - \vartheta_v = (x_1 - \vartheta, x_2 - \vartheta, \dots, x_n - \vartheta).$$

Da  $\hat{\vartheta}$  eine äquivariante Schätzfunktion ist, so gilt  $\hat{\vartheta}(x) - \vartheta = \hat{\vartheta}(x - \vartheta_v)$ , und daher können wir schreiben

$$\begin{aligned} \text{MSE}(\hat{\vartheta}(X)) &= E_{\vartheta}(\hat{\vartheta}(X) - \vartheta)^2 = \int_{\mathfrak{X}} (\hat{\vartheta}(x) - \vartheta)^2 f_{\vartheta}^{(n)}(x) dx \\ &= \int_{\mathfrak{X}} \hat{\vartheta}^2(x - \vartheta_{\mathbf{v}}) f^{(n)}(x - \vartheta_{\mathbf{v}}) dx = \int_{\mathfrak{X}} \hat{\vartheta}^2(x) f^{(n)}(x) dx = E_0(\hat{\vartheta}^2). \end{aligned}$$

$\hat{\vartheta}$  ist äquivariant und somit darstellbar als

$$\hat{\vartheta} = \hat{\vartheta}(X) = e(X) + h(X),$$

wobei  $e: \mathfrak{X} \rightarrow \mathbb{R}$  eine beliebige äquivariante Statistik und  $h: \mathfrak{X} \rightarrow \mathbb{R}$  eine beliebige invariante Statistik ist. Nun sei  $T: \mathfrak{X} \rightarrow \mathfrak{X}$  eine beliebige Äquivalenzklassen-Statistik, und wir wollen den Satz über den bedingten Erwartungswert anwenden auf die beiden Zufallsgrößen  $\hat{\vartheta}^2(X)$  und  $T(X)$ :

$$\begin{aligned} \text{MSE}(\hat{\vartheta}(X)) &= E_0(\hat{\vartheta}^2) = E_0(E_0(\hat{\vartheta}^2|T)) = \int_{\mathbb{R}^{n-1}} E_0(\hat{\vartheta}^2|T=t) g(t) dt \\ (1) \quad &= \int_{\mathbb{R}^{n-1}} E_0([e(X) + h(t)]^2|T=t) g(t) dt \rightarrow \min, \end{aligned}$$

wobei  $g(t)$  die Dichte der  $(n-1)$ -dimensionalen Zufallsgröße  $T(X)$  bezeichnet. Aufgrund der Minimumseigenschaft des Erwartungswerts gilt

$$E_0([e(X) + h(t)]^2|T=t) = \min$$

falls

$$(2) \quad h(t) = -E_0(e(X)|T=t).$$

Wenn wir also  $h(t)$  definieren gemäß (2) für alle Werte der Statistik  $T$ , dann wird der Integrand im letzten Integral von (1) punktweise minimal, und damit ist die optimale Schätzfunktion gegeben durch

$$(3) \quad \hat{\vartheta}(X) = e(X) + h(T(X)) = e(X) - E_0(e(X)|T(X)).$$

Dabei ist  $e: \mathfrak{X} \rightarrow \mathbb{R}$  eine beliebige äquivariante Statistik und  $T: \mathfrak{X} \rightarrow \mathfrak{X}$  eine beliebige Äquivalenzklassen-Statistik.

*Folgerung:*

Die optimale Schätzfunktion  $\hat{\vartheta}$  gemäß (3) ist stets *erwartungstreu*, denn es gilt

$$(4) \quad E_{\vartheta}(\hat{\vartheta}) = E_{\vartheta}(e(X)) + E_{\vartheta}(h(T(X))).$$

Nun ist  $e: \mathfrak{X} \rightarrow \mathbb{R}$  äquivariant ist, und somit gilt

$$E_{\vartheta}(e(X)) = E_0(e(X)) + \vartheta;$$

weiter ist  $h(T): \mathfrak{X} \rightarrow \mathbb{R}$  invariant, und somit gilt

$$E_{\vartheta}[h(T(X))] = E_0[h(T(X))] = E_0[-E_0(e(X)|T(X))] = -E_0(e(X)),$$

und damit gilt

$$E_{\vartheta}(\hat{\vartheta}) = E_0(e(X)) + \vartheta - E_0(e(X)) = \vartheta,$$

was zu beweisen war.

*Beispiel:*

$X_1, \dots, X_n$  seien unabhängige Zufallsgrößen je mit einer Normalverteilung  $N(\mu, \sigma^2)$ . Bei festem  $\sigma^2$  bildet die Familie  $\{N(\mu, \sigma^2) | \mu \in \mathbb{R}\}$  ein Lagemodell, und wir suchen den optimalen invarianten Schätzer für  $\mu$ . Gemäß (3) hat dieser Schätzer die Form

$$\hat{\mu}(X) = e(X) + h(T(X)),$$

wobei

$$h(t) = -E_0(e(X) | T(X) = t).$$

Wir wählen

$$T(X) = (0, X_2 - X_1, X_3 - X_1, \dots, X_n - X_1);$$

$$e(X) = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n).$$

Dann gilt für

$$E(\bar{X}(X_j - X_1)) = 0 \quad \text{für } j = 2, 3, \dots, n;$$

somit sind die beiden Zufallsgrößen  $\bar{X}$  und  $T(X)$  unkorreliert und daher auch stochastisch unabhängig (multivariate Normalverteilung!). Daher gilt

$$h(T(X)) = -E_0(\bar{X} | T(X) = t) = -E_0(\bar{X}) = 0,$$

und wir erhalten

$$\hat{\mu}(X) = e(X) + h(T(X)) = e(X) = \bar{X}.$$

Die Schätzfunktion  $\hat{\mu} = \bar{X}$  ist also invariant und erwartungstreu, und es gilt  $\text{MSE}(\hat{\mu}) = \min$ . Dies gilt für beliebige Werte von  $\sigma^2$ .

#### 4. Lösung der Optimalitätsaufgabe, Teil 2

Es sei

$$\begin{aligned} X &= (X_1, \dots, X_n) \text{ unser Zufallsvektor;} \\ f_X(x_1, \dots, x_n) &\text{ die zugehörige Dichtefunktion;} \\ T(X) &= (0, T_2, T_3, \dots, T_n) \text{ mit } T_j = X_j - X_1; \\ e(X) &= X_1; \\ h(t) &= E_0(X_1 | T=t). \end{aligned}$$

Wir wollen den Erwartungswert  $h(t)$  berechnen. Es sei

$$Y = (Y_1, Y_2, \dots, Y_n) = (X_1, X_2 - X_1, \dots, X_n - X_1) = (X_1, T_2, \dots, T_n),$$

und  $f_Y(y_1, \dots, y_n)$  sei die Dichte von  $Y$ . Wir wollen die Dichte  $f_Y$  bestimmen. Es sei

$$A = \left( \begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ -1 & 1 & & \mathbf{0} \\ \vdots & & \ddots & \\ -1 & \mathbf{0} & & 1 \end{array} \right); \text{ dann ist } A^{-1} = \left( \begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ 1 & 1 & & \mathbf{0} \\ \vdots & & \ddots & \\ 1 & \mathbf{0} & & 1 \end{array} \right),$$

und es ist  $\det(\mathbf{A}) = \det(\mathbf{A}^{-1}) = 1$ . Wenn wir die obigen Zufallsvektoren  $X$  und  $Y$  als Spaltenvektoren auffassen, dann gilt  $Y = \mathbf{A}X$ , und die Funktionaldeterminante der linearen Transformation ist 1.

Somit ist die Dichte von  $Y$  gegeben durch

$$f_Y(y) = f_X(\mathbf{A}^{-1}y) = f_X(y_1, y_1 + y_2, y_1 + y_3, \dots, y_1 + y_n).$$

Wir bezeichnen die gemeinsame Verteilung von  $X_1$  und  $T$  mit  $f_{(X_1, T)}$  und die bedingte Dichte von  $X_1$  gegeben  $T = t$  mit  $f_{X_1|T=t}$ . Dann gilt

$$f_{X_1|T=t}(x_1) = \frac{f_{(X_1, T)}(x_1, t)}{f_T(t)} = \frac{f_Y(x_1, t_2, \dots, t_n)}{f_T(t)}.$$

Die Randdichte von  $T$  ist gegeben durch

$$\begin{aligned} f_T(t) &= \int_{\mathbb{R}} f_Y(y_1, t_2, \dots, t_n) dy_1 \\ &= \int_{\mathbb{R}} f_X(y_1, t_2 + y_1, \dots, t_n + y_1) dy_1 \end{aligned}$$

Mit der Substitution  $z = x_1 - y_1$  erhalten wir

$$z = x_1 - y_1 \text{ und weiter } t_j + y_1 = (x_j - x_1) + (x_1 - z) = x_j - z,$$

und somit können wir schreiben

$$f_T(t) = \int_{\mathbb{R}} f_X(x_1 - z, x_2 - z, \dots, x_n - z) dz.$$

Jetzt wollen wir den Erwartungswert  $h(t)$  berechnen. Es ist

$$\begin{aligned} h(t) &= E_0(X_1 | T=t) = \int_{\mathbb{R}} u f_{X_1|T=t}(u) du \\ &= \frac{1}{f_T(t)} \int_{\mathbb{R}} u f_Y(u, t_2, \dots, t_n) du \\ &= \frac{1}{f_T(t)} \int_{\mathbb{R}} u f_X(u, t_2 + u, \dots, t_n + u) du. \end{aligned}$$

Mit der Substitution  $z = x_1 - y_1$  (wie oben) erhalten wir

$$\begin{aligned} h(t) &= \frac{1}{f_T(t)} \int_{\mathbb{R}} (x_1 - z) f_X(x_1 - z, x_2 - z, \dots, x_n - z) dz \\ &= x_1 - \frac{1}{f_T(t)} \int_{\mathbb{R}} z f_X(x_1 - z, x_2 - z, \dots, x_n - z) dz, \end{aligned}$$

und somit ergibt sich für unsere optimale Schätzfunktion die Darstellung

$$\begin{aligned} \hat{\vartheta}(x) &= e(x) - h(t) = x_1 - h(t) \\ &= \frac{\int_{\mathbb{R}} z f_X(x_1 - z, x_2 - z, \dots, x_n - z) dz}{\int_{\mathbb{R}} f_X(x_1 - z, x_2 - z, \dots, x_n - z) dz} \end{aligned}$$

Wegen der Unabhängigkeit der Komponenten  $X_1, \dots, X_n$  können wir schließlich schreiben

$$(5) \quad \hat{\vartheta}(x) = \hat{\vartheta}(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz}{\int_{\mathbb{R}} f(x_1 - z) \cdots f(x_n - z) dz}.$$

Diese Darstellung geht zurück auf Pitman, und daher nennt man den Schätzer  $\hat{\vartheta}$  auch einen Pitman-Schätzer.

*Bemerkungen:*

a) Die Voraussetzungen für die Optimalität des Pitman-Schätzers sind minimal; die Komponenten  $X_1, \dots, X_n$  müssen eine Dichte besitzen, und die Varianz von  $\hat{\vartheta}$  muß endlich sein.

b) Die Likelihood-Funktion für den Parameter  $\vartheta$  ist

$$L(\vartheta) = f_{\vartheta}(x_1) \cdots f_{\vartheta}(x_n) = f(x_1 - \vartheta) \cdots f(x_n - \vartheta).$$

Nun betrachten wir die normierte Likelihood-Funktion als Dichtefunktion:

$$g(\vartheta) = cL(\vartheta) \quad \text{mit} \quad \frac{1}{c} = \int_{-\infty}^{\infty} L(\vartheta) d\vartheta = \int_{\mathbb{R}} f(x_1 - \vartheta) \cdots f(x_n - \vartheta) d\vartheta.$$

Der Erwartungswert der Dichte  $g$  ist

$$\int_{\mathbb{R}} \vartheta g(\vartheta) d\vartheta = \frac{1}{c} \int_{\mathbb{R}} \vartheta f(x_1 - \vartheta) \cdots f(x_n - \vartheta) d\vartheta = \hat{\vartheta}(x_1, \dots, x_n).$$

Der Erwartungswert der normierten Likelihood-Funktion ist also gerade der Pitman-Schätzer. Der Modalwert der Likelihood-Funktion (Wert von  $\vartheta$ , für den die Likelihood-Funktion maximal wird), ist der Maximum-Likelihood-Schätzer. Der Pitman-Schätzer ist also eng verwandt mit dem Maximum-Likelihood-Schätzer.

c) Im Spezialfall  $n = 1$  erhalten wir

$$\hat{\vartheta}(x) = \hat{\vartheta}(x_1) = \frac{\int_{\mathbb{R}} z f(x_1 - z) dz}{\int_{\mathbb{R}} f(x_1 - z) dz}.$$

Da der Nenner ist 1, so haben wir

$$(6) \quad \hat{\vartheta}(x_1) = \int_{\mathbb{R}} z f(x_1 - z) dz = \int_{\mathbb{R}} (x_1 - t) f(t) dt = x_1 - \int_{\mathbb{R}} t f(t) dt.$$

Offensichtlich gilt für  $\vartheta = 0$

$$E_0(\hat{\vartheta}) = E_0(X_1) - \int_{\mathbb{R}} tf(t)dt = 0,$$

und damit bestätigt sich, daß der Pitman-Schätzer erwartungstreu ist. Für  $n = 1$  existiert der Pitman-Schätzer nur, falls der Erwartungswert von  $E_0(X_1)$  existiert.

c) Im Spezialfall  $n = 2$  erhalten wir

$$\hat{\vartheta}(x) = \hat{\vartheta}(x_1, x_2) = \frac{\int z f(x_1 - z) f(x_2 - z) dz}{\int f(x_1 - z) f(x_2 - z) dz}.$$

Wir betrachten die Dichte

$$g(z) = c f(x_1 - z) f(x_2 - z) \quad \text{mit} \quad \frac{1}{c} = \int_{\mathbb{R}} f(x_1 - z) f(x_2 - z) dz.$$

Es gilt dann

$$(7) \quad \hat{\vartheta}(x) = \hat{\vartheta}(x_1, x_2) = \int z g(z) dz.$$

Falls die Dichte  $f$  symmetrisch ist um 0, so ist die Dichte  $g$  symmetrisch um  $\frac{1}{2}(x_1 + x_2)$ . In diesem Fall ist also  $\hat{\vartheta}(x_1, x_2) = \frac{1}{2}(x_1 + x_2)$ .

## 5. Beispiele

### 1. Beispiel

$X_1, \dots, X_n$  seien unabhängige Zufallsgrößen je mit einer Normalverteilung  $N(\mu, \sigma^2)$ . Bei festem  $\sigma^2$  betrachten wir das Lagemodell  $\{N(\mu, \sigma^2) | -\infty < \mu < \infty\}$ . Dann ist der optimale Schätzer für  $\mu$  gegeben durch

$$(8) \quad \hat{\mu}(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz}{\int_{\mathbb{R}} f(x_1 - z) \cdots f(x_n - z) dz}, \quad \text{wobei} \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Der Zähler von (8) kann geschrieben werden als

$$\text{Zähler} = \int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \int_{\mathbb{R}} z \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - z)^2\right) dz.$$

Nun gilt

$$\sum (x_i - z)^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - z)^2,$$

und daher ist

$$\text{Zähler} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \bar{x})^2\right) \underbrace{\int_{\mathbb{R}} z \exp\left(-\frac{n}{2\sigma^2} (z - \bar{x})^2\right) dz}_{=\bar{x}\sigma\sqrt{2\pi/n}}.$$

Auf analoge Weise kann der Nenner von (8) geschrieben werden als

$$\text{Nenner} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \bar{x})^2\right) \underbrace{\int_{\mathbb{R}} \exp\left(-\frac{n}{2\sigma^2} (z - \bar{x})^2\right) dz}_{=\sigma\sqrt{2\pi/n}},$$

und somit erhalten wir

$$\hat{\mu}(x_1, \dots, x_n) = \frac{\text{Zähler}}{\text{Nenner}} = \bar{x}.$$

Das arithmetische Mittel  $\hat{\mu} = \bar{X}$  ist also invariant und erwartungstreu, und es ist optimal bezüglich des mittleren quadratischen Fehlers unter allen invarianten Schätzfunktionen:  $\text{MSE}(\hat{\vartheta}) = \min$ . Dies gilt für beliebige Werte von  $\sigma^2$ . Damit bestätigt sich ein Ergebnis, das wir bereits oben gefunden haben.

## 2. Beispiel

$X_1, \dots, X_n$  seien unabhängige Zufallsgrößen je mit einer Gleichverteilung (Rechteckverteilung) im Intervalle  $(\vartheta-1, \vartheta+1)$  mit  $-\infty < \vartheta < \infty$ . Die Familie der zulässigen Verteilungen bildet wiederum ein Lagemodell, und die Dichte der einzelnen Komponenten  $X_j$  ist gegeben durch

$$f_{\vartheta}(x) = f(x - \vartheta), \text{ wobei } f(x) = \begin{cases} \frac{1}{2} & \text{falls } |x| < 1 \\ 0 & \text{sonst.} \end{cases}$$

Der optimale Schätzer für  $\vartheta$  ist gegeben durch

$$(9) \quad \hat{\vartheta}(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz}{\int_{\mathbb{R}} f(x_1 - z) \cdots f(x_n - z) dz}.$$

Das Produkt  $f(x_1 - z) \cdots f(x_n - z)$  ist nur dann ungleich Null, wenn

$$-1 < x_j - z < 1 \quad \forall j, \text{ d.h. wenn } z - 1 < x_j < z + 1 \quad \forall j;$$

daraus folgt  $z < m + 1$  und  $z > M - 1$ , wobei  $m = \min(x_1, \dots, x_n)$  und  $M = \max(x_1, \dots, x_n)$ . Wir können daher den Zähler in (9) schreiben als

$$\begin{aligned} \text{Zähler} &= \int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz = \int_{M-1}^{m+1} z f(x_1 - z) \cdots f(x_n - z) dz = \\ &= \left(\frac{1}{2}\right)^n \int_{M-1}^{m+1} z dz = \left(\frac{1}{2}\right)^{n+1} \left[ (m+1)^2 - (M-1)^2 \right]. \end{aligned}$$

Auf analoge Weise gilt

$$\text{Nenner} = \int_{M-1}^{m+1} f(x_1 - z) \cdots f(x_n - z) dz = \left(\frac{1}{2}\right)^n \int_{M-1}^{m+1} dz = \left(\frac{1}{2}\right)^n [(m+1) - (M-1)].$$

Somit erhalten wir das Ergebnis

$$\hat{\vartheta}(x_1, \dots, x_n) = \frac{\text{Zähler}}{\text{Nenner}} = \frac{1}{2} \frac{(m-1)^2 - (M+1)^2}{(m-1) - (M+1)} = \frac{1}{2}(m+M).$$

Wir wissen bereits, daß das Paar  $(m, M)$  eine suffiziente Statistik für  $\vartheta$  ist, und daß die Konvergenzgeschwindigkeit beim Schätzer  $\hat{\vartheta}$  gegeben ist durch den Faktor  $n$  anstelle des üblichen Faktors  $\sqrt{n}$ . Jetzt wissen wir, daß  $\hat{\vartheta}$  optimal ist bezüglich des mittleren quadratischen Fehlers unter allen invarianten Schätzfunktionen:  $\text{MSE}(\hat{\vartheta}) = \min$ . Die Likelihood-Funktion  $L(\vartheta) = f(x_1 - \vartheta) \cdots f(x_n - \vartheta)$  kann nur die Werte 0 und  $(\frac{1}{2})^n$  annehmen; sie ist ungleich 0 und damit maximal, falls  $m + 1 < \vartheta < M - 1$ . Daher ist jeder Wert in diesem Intervall ein Maximum-Likelihood-Schätzer für  $\vartheta$ ; der Mittelpunkt dieses Intervalls ist der Pitman-Schätzer, und somit ist in unserem Beispiel der Pitman-Schätzer stets auch ein Maximum-Likelihood-Schätzer.

### 3. Beispiel

$X_1, \dots, X_n$  seien unabhängige Zufallsgrößen je mit einer Dreiecksverteilung im Intervall  $(\vartheta-1, \vartheta+1)$  mit  $-\infty < \vartheta < \infty$ . Die Familie der zulässigen Verteilungen bildet wiederum ein Lagemodell, und die Dichte der einzelnen Komponenten  $X_i$  ist gegeben durch

$$f_{\vartheta}(x) = f(x - \vartheta), \text{ wobei } f(x) = \begin{cases} 0 & \text{falls } x \leq -1 \\ x+1 & \text{falls } -1 < x \leq 0 \\ 1-x & \text{falls } 0 < x \leq 1 \\ 0 & \text{falls } x > 1. \end{cases}$$

Der optimale Schätzer für  $\vartheta$  ist gegeben durch

$$(10) \hat{\vartheta}(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz}{\int_{\mathbb{R}} f(x_1 - z) \cdots f(x_n - z) dz}.$$

Wie im Beispiel 2 ist das Produkt  $f(x_1 - z) \cdots f(x_n - z)$  ist nur dann ungleich Null, wenn  $M - 1 < z < m + 1$ , wobei  $m = \min(x_1, \dots, x_n)$  und  $M = \max(x_1, \dots, x_n)$ . Wir können daher den Zähler in (10) schreiben als

$$\text{Zähler} = \int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz = \int_{M-1}^{m+1} z f(x_1 - z) \cdots f(x_n - z) dz.$$

Analoges gilt für den Nenner. Eine weitere Vereinfachung scheint im allgemeinen Fall nicht möglich zu sein; es existiert ja auch keine nicht-triviale suffiziente Statistik für  $\vartheta$ . Für  $n=1$  gilt  $\hat{\vartheta}(x_1) = x_1$ , und für  $n=2$  ist  $\hat{\vartheta}(x_1, x_2) = \frac{1}{2}(x_1 + x_2)$  aufgrund der Symmetrie der Dichte  $f$ . Tabelle 1 gibt für verschiedene Stichproben vom Umfang  $n=3$  die Schätzwerte von vier verschiedenen Schätzfunktionen an. Es zeigt sich, daß der Pitman-Schätzer im allgemeinen verschieden ist vom Maximum-Likelihood-Schätzer. Da der Pitman-Schätzer optimal ist bezüglich des mittleren quadratischen Fehlers, und da der Schätzer stets erwartungstreu ist, so besitzt der Pitman-Schätzer unter den vier Schätzfunktionen die kleinste Varianz. Es ist zu vermuten, daß auch der ML-Schätzer asymptotisch optimal ist und die gleiche asymptotische Verteilung besitzt wie der Pitman-Schätzer. Die asymptotische Verteilung des Pitman-Schätzers und die Konvergenzgeschwindigkeit sind mir nicht bekannt.

**Tabelle 1:** Schätzwerte für verschiedene Stichproben vom Umfang  $n=3$ :  
(Pi und ML berechnet mit Maple-Arbeitsblatt *pitman.mws*)

$x_1$	$x_2$	$x_3$	Pi	ML	$\bar{x}$	Ze
1	2	2	1.6	1.667	1.667	2
1.1	2	2.1	1.682	1.751	1.733	2
1.2	2	2.2	1.746	1.837	1.8	2
1.3	2	2.3	1.845	1.926	1.867	2
1.4	2	2.4	1.923	2	1.933	2
1.5	2	2.5	2	2	2	2
1.6	2	2.6	2.077	2	2.067	2
1.7	2	2.7	2.155	2.074	2.133	2
1.8	2	2.8	2.236	2.162	2.2	2
1.9	2	2.9	2.318	2.249	2.267	2
2	2	3	2.4	2.333	2.333	2

Pi = Pitman-Schätzer,  
ML = Maximum-Likelihood-Schätzer,  
 $\bar{x}$  = arithmetisches Mittel,  
Ze = Zentralwert

### 4. Beispiel

$X_1, \dots, X_n$  seien unabhängige Zufallsgrößen je mit einer Cauchyverteilung mit der Dichte

$$f_{\vartheta}(x) = f(x - \vartheta), \text{ wobei } f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Die Familie der zulässigen Verteilungen bildet wiederum ein Lagemodell. Der optimale Schätzer für  $\vartheta$  ist gegeben durch

$$\hat{\vartheta}(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz}{\int_{\mathbb{R}} f(x_1 - z) \cdots f(x_n - z) dz}.$$

Da keine nicht-triviale suffiziente Statistik für  $\vartheta$  existiert, so läßt sich diese Darstellung nicht weiter vereinfachen. Für  $n = 1$  gilt aufgrund von (6)

$$\hat{\vartheta}(x_1) = \int_{\mathbb{R}} z f(x_1 - z) dz = \int_{\mathbb{R}} (x_1 - t) f(t) dt = x_1 - \int_{\mathbb{R}} t f(t) dt.$$

Das letzte Integral existiert jedoch nicht, und daher existiert auch der Pitman-Schätzer nicht im Falle  $n = 1$ . Für  $n = 2$  ist

$$(11) \quad \hat{\vartheta}(x_1, x_2) = \frac{\int z f(x_1 - z) f(x_2 - z) dz}{\int f(x_1 - z) f(x_2 - z) dz}.$$

Es gilt

$$z f(x_1 - z) f(x_2 - z) = \left(\frac{1}{\pi}\right)^2 \frac{z}{\left[1 + (x_1 - z)^2\right] \left[1 + (x_2 - z)^2\right]} \sim \left(\frac{1}{\pi}\right)^2 \frac{1}{z^3} \text{ für } z \rightarrow \infty,$$

und somit existieren die beiden Integrale in (11). Wegen der Symmetrie der Dichte  $f$  ist also

$$\hat{\vartheta}(x_1, x_2) = \frac{1}{2}(x_1 + x_2).$$

Tabelle 2 gibt für verschiedene Stichproben vom Umfang  $n = 3$  die Schätzwerte von vier verschiedenen Schätzfunktionen an. Es zeigt sich auch hier, daß der Pitman-Schätzer im allgemeinen verschieden ist vom Maximum-Likelihood-Schätzer. Da der Pitman-Schätzer optimal ist bezüglich des mittleren quadratischen Fehlers, und da der Schätzer stets erwartungstreu ist, so besitzt der Pitman-Schätzer unter den vier Schätzfunktionen die kleinste Varianz. Es ist zu vermuten, daß auch der ML-Schätzer asymptotisch optimal ist und die gleiche asymptotische Verteilung besitzt wie der Pitman-Schätzer. Die asymptotische Verteilung des Pitman-Schätzers ist mir nicht bekannt.

**Tabelle 2:** Schätzwerte für verschiedene Stichproben vom Umfang  $n = 3$ :  
(Pi und ML berechnet mit Maple-Arbeitsblatt *pitman.mws*)

$x_1$	$x_2$	$x_3$	Pi	ML	$\bar{x}$	Ze
0	2	2	1.5	1.775	1.333	2
0.5	2	2.5	1.763	1.954	1.667	2
1	2	3	2	2	2	2
1.5	2	3.5	2.238	2.046	2.333	2
2	2	4	2.5	2.225	2.667	2

Pi = Pitman-Schätzer,  
ML = Maximum-Likelihood-Schätzer,  
 $\bar{x}$  = arithmetisches Mittel,  
Ze = Zentralwert

## 5. Beispiel

$X_1, \dots, X_n$  seien unabhängige Zufallsgrößen je mit einer zweiseitigen Exponentialverteilung mit der Dichte

$$f_{\vartheta}(x) = f(x - \vartheta), \text{ wobei } f(x) = \frac{1}{2}e^{-|x|}, x \in \mathbb{R}.$$

Die Familie der zulässigen Verteilungen bildet wiederum ein Lagemodell. Der optimale Schätzer für  $\vartheta$  ist gegeben durch

$$\hat{\vartheta}(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z f(x_1 - z) \cdots f(x_n - z) dz}{\int_{\mathbb{R}} f(x_1 - z) \cdots f(x_n - z) dz}.$$

Da keine nicht-triviale suffiziente Statistik für  $\vartheta$  existiert, so läßt sich diese Darstellung nicht weiter vereinfachen. Für  $n = 1$  gilt aufgrund von (6)

$$\hat{\vartheta}(x_1) = \int_{\mathbb{R}} z f(x_1 - z) dz = \int_{\mathbb{R}} (x_1 - t) f(t) dt = x_1 - \int_{\mathbb{R}} t f(t) dt = x_1,$$

da die Dichte  $f$  symmetrisch ist bezüglich null. Für  $n = 2$  ist aufgrund von (7)

$$\hat{\vartheta}(x_1, x_2) = \frac{\int_{\mathbb{R}} z f(x_1 - z) f(x_2 - z) dz}{\int_{\mathbb{R}} f(x_1 - z) f(x_2 - z) dz} = \frac{1}{2}(x_1 + x_2).$$

Tabelle 3 gibt für verschiedene Stichproben vom Umfang  $n = 3$  die Schätzwerte von vier verschiedenen Schätzfunktionen an. Es zeigt sich auch hier, daß der Pitman-Schätzer im allgemeinen verschieden ist vom Maximum-Likelihood-Schätzer. Da der Pitman-Schätzer optimal ist bezüglich des mittleren quadratischen Fehlers, und da der Schätzer stets erwartungstreu ist, so besitzt der Pitman-Schätzer unter den vier Schätzfunktionen die kleinste Varianz. Aufgrund der Minimumeigenschaft des Zentralwerts ist der Maximum-Likelihood-Schätzer für beliebiges  $n$  identisch zum Zentralwert  $Ze(x_1, \dots, x_n)$ . Es ist zu vermuten, daß auch der ML-Schätzer asymptotisch optimal ist und die gleiche asymptotische Verteilung besitzt wie der Pitman-Schätzer. Die asymptotische Verteilung des Pitman-Schätzers ist mir nicht bekannt.

**Tabelle 3:** Schätzwerte für verschiedene Stichproben vom Umfang  $n = 3$ :  
(Pi berechnet mit Maple-Arbeitsblatt *pitman.mws*)

$x_1$	$x_2$	$x_3$	Pi	ML	$\bar{x}$	Ze
0	2	2	1.527	2	1.333	2
0.5	2	2.5	1.779	2	1.667	2
1	2	3	2	2	2	2
1.5	2	3.5	2.221	2	2.333	2
2	2	4	2.473	2	2.667	2

Pi = Pitman-Schätzer,  
ML = Maximum-Likelihood-Schätzer,  
 $\bar{x}$  = arithmetisches Mittel,  
Ze = Zentralwert

## 18. Zur Bayes'schen Statistik

### 1. Binomialverteilung mit Beta-Verteilung als a-priori-Verteilung

#### 1.1 Ausgangssituation

Es sei

$$X \sim \text{Bi}(n, \vartheta);$$

$n$  bekannt;

$\vartheta$  unbekannt.

Der unbekannte Parameter  $\vartheta$  sei eine Realisation einer Zufallsgröße  $\vartheta_{\text{rv}}$  (rv = random variable) mit einer Beta-Verteilung:

$$\vartheta_{\text{rv}} \sim \text{Be}(a, b), \quad a, b > 0, \quad a, b \text{ bekannt};$$

$$\text{Dichte von } \vartheta_{\text{rv}}: f(\vartheta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \vartheta^{a-1} (1-\vartheta)^{b-1}, \quad 0 < \vartheta < 1;$$

$$E(\vartheta_{\text{rv}}) = \frac{a}{a+b},$$

$$\text{Var}(\vartheta_{\text{rv}}) = \frac{ab}{(a+b)^2(a+b+1)},$$

$$\text{Mo}(\vartheta_{\text{rv}}) = \frac{a-1}{a+b-2} \text{ falls } a, b > 1.$$

$$\text{Falls } a = b = 1: \vartheta_{\text{rv}} \sim \text{R}(0, 1).$$

*Zweistufiges Experiment:*

$$1. \text{ Stufe: } \vartheta_{\text{rv}} \text{ realisieren gemäß } \text{Be}(a, b) \quad \rightarrow \quad \vartheta \in (0, 1);$$

$$2. \text{ Stufe: } X \text{ realisieren gemäß } \text{Bi}(n, \vartheta) \quad \rightarrow \quad k \in \{0, 1, \dots, n\}.$$

*Bayes'sche Statistik:*

$a, b, n$  bekannte Parameter ( $a, b$  sogenannte Hyperparameter);

$\vartheta$  unbekannte Realisation von  $\vartheta_{\text{rv}}$ ;

$k$  vorliegende Beobachtung von  $X$ .

Ziel:  $\vartheta$  schätzen; Vertrauensgrenzen für  $\vartheta$  bestimmen.

*A-priori-Verteilung* von  $\vartheta_{\text{rv}}$ :  $\text{Be}(a, b)$ ;

*a-posteriori-Verteilung* von  $\vartheta_{\text{rv}}$ : Bedingte Verteilung von  $\vartheta_{\text{rv}}$  gegeben  $X = k$ .

#### 1.2 A-posteriori-Verteilung von $\vartheta_{\text{rv}}$

Wir betrachten also die folgenden Verteilungen:

$$\vartheta_{\text{rv}} \sim \text{Be}(a, b) \quad \text{stetige Verteilung;}$$

$$X \sim \text{Bi}(n, \vartheta) \quad \text{diskrete Verteilung;}$$

$$\vartheta_{\text{rv}} | X=k \quad \text{stetige Verteilung.}$$

Die Dichte von  $\vartheta_{\text{rv}} | X=k$  ist gegeben durch

$$(1) \quad f_k(\vartheta) = \frac{f(\vartheta)p_\vartheta(k)}{P\{X=k\}} = \frac{1}{P\{X=k\}} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \vartheta^{a-1}(1-\vartheta)^{b-1} \times \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}$$

$$= \frac{1}{P\{X=k\}} \times \binom{n}{k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \vartheta^{a+k-1}(1-\vartheta)^{b+n-k-1}$$

Nun ist

$$(2) \quad P\{X=k\} = P\{\vartheta_{\text{rv}} = \text{beliebiger Wert}, X=k\} = \int_0^1 f(\vartheta)p_\vartheta(k)d\vartheta$$

$$= \binom{n}{k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \vartheta^{a+k-1}(1-\vartheta)^{b+n-k-1} d\vartheta$$

$$= \binom{n}{k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)};$$

somit ergibt sich aus (1) und (2)

$$f_k(\vartheta) = \frac{\Gamma(a+b+n)}{\Gamma(a+k)\Gamma(b+n-k)} \vartheta^{a+k-1}(1-\vartheta)^{b+n-k-1},$$

und dies ist wiederum die Dichte einer Beta-Verteilung. Wir haben also das Ergebnis

$$\frac{\vartheta_{\text{rv}} | X=k}{= Y} \sim \text{Be}(a+k, b+n-k);$$

$$E(Y) = \frac{a+k}{a+b+n} \quad (\approx p = \frac{k}{n}, \text{ falls } n \text{ und } k \text{ groß});$$

$$\text{Var}(Y) = \frac{(a+k)(b+n-k)}{(a+b+n)^2(a+b+n+1)} \quad (\approx \frac{p(1-p)}{n}, \text{ falls } n \text{ und } k \text{ groß});$$

$$\text{Mo}(Y) = \frac{a+k-1}{a+b+n-2} \quad (\approx p, \text{ falls } n \text{ und } k \text{ groß}).$$

*A-posteriori-Modus-Schätzung* (=Maximum-Likelihood-Schätzung in der Bayes-Theorie):

$$\hat{\vartheta} = \text{Mo}(\vartheta_{\text{rv}} | X=k) = \frac{a+k-1}{a+b+n-2} \quad (\approx p \text{ falls } n \text{ und } k \text{ groß}).$$

Vertrauensgrenzen für  $\vartheta$ :

$$\left. \begin{array}{l} \vartheta_\ell = \text{linkes} \\ \vartheta_r = \text{rechtes} \end{array} \right\} \alpha\text{-Quantil der a-posteriori-Verteilung } \text{Be}(a+k, b+n-k), \text{ d.h.}$$

$$\vartheta_\ell = B_\alpha(a+k, b+n-k);$$

$$\vartheta_r = B_{1-\alpha}(a+k, b+n-k).$$

$$\text{Es gilt dann: } P\{\vartheta_\ell < \vartheta_{\text{rv}} | X=k < \vartheta_r\} = 1 - 2\alpha.$$

Im Bayes-Modell wird die Realisation  $k$  der Zufallsgröße  $X$  also als fest gegeben betrachtet; die Vertrauensgrenzen  $\vartheta_\ell$  und  $\vartheta_r$  sind dann feste Zahlen, und die Wahrscheinlichkeitsaussage bezieht sich auf den zufälligen Parameter  $\vartheta_{\text{rv}} | X=k$ , d.h. auf die a-posteriori-Verteilung von  $\vartheta_{\text{rv}}$ .

Im Spezialfall  $a = b = 1$  gilt:

$$\vartheta_{\text{rv}} \sim \text{R}(0,1) \quad (\text{nicht-informative a-priori-Verteilung});$$

$$\vartheta_{\text{rv}} | X=k \sim \text{Be}(k+1, n-k+1);$$

$$\hat{\vartheta} = \text{Mo}(\vartheta_{\text{rv}} | X=k) = \frac{k}{n} = p \quad (\text{wie im klassischen Fall}).$$

$$\vartheta_{\ell} = B_{\alpha}(k+1, n-k+1)$$

$$\vartheta_r = B_{1-\alpha}(k+1, n-k+1)$$

Klassische Vertrauensgrenzen für  $\vartheta$  (mit Konfidenzinterpretation):

$$\vartheta_{\ell} = B_{\alpha}(k, n-k+1) \quad \text{falls } k > 0 \quad (\vartheta_{\ell} = 0 \text{ falls } k = 0);$$

$$\vartheta_r = B_{1-\alpha}(k+1, n-k) \quad \text{falls } k < n \quad (\vartheta_r = 1 \text{ falls } k = n).$$

*Bemerkung:*

Im Spezialfall  $a = b = 1$  stimmt die Bayes-Schätzung  $\hat{\vartheta}$  mit der klassischen Maximum-Likelihood-Schätzung überein, und die Vertrauensgrenzen nach Bayes sind stets etwas enger als die klassischen Vertrauensgrenzen. Die klassische Konfidenzinterpretation ist jedoch korrekt für beliebige Werte des unbekanntes Parameters  $\vartheta$  und somit auch bei Vorliegen einer *beliebigen* a-priori-Verteilung für  $\vartheta$ .

### 1.3 Randverteilung von $X$ ; Rechnen mit bedingten Erwartungswerten

Wir betrachten das zweistufige Bayes-Experiment (wie oben) und wollen die Randverteilung (Marginalverteilung) von  $X$  untersuchen. Gemäß (2) gilt:

$$\begin{aligned} P\{X = k\} &= P\{\vartheta_{\text{rv}} = \text{beliebiger Wert}, X = k\} = \int_0^1 f(\vartheta) p_{\vartheta}(k) d\vartheta = \\ &= \binom{n}{k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)}. \end{aligned}$$

Nun wollen wir den Erwartungswert und die Varianz von  $X$  berechnen mit Hilfe der Sätze über den bedingten Erwartungswert und die bedingte Varianz. Der Satz vom bedingten Erwartungswert lautet:

$$E(X) = E\left(E(X | \vartheta_{\text{rv}})\right).$$

Nun ist

$$E(X | \vartheta_{\text{rv}} = \vartheta) = n\vartheta,$$

$$E(X | \vartheta_{\text{rv}}) = n\vartheta_{\text{rv}} \quad (\text{bedingter Erwartungswert als Zufallsgröße}),$$

und somit

$$E(X) = E\left(E(X | \vartheta_{\text{rv}})\right) = E(n\vartheta_{\text{rv}}) = nE(\vartheta_{\text{rv}}) = n \frac{a}{a+b}.$$

Der Satz von der bedingten Varianz lautet:

$$(3) \quad \text{Var}(X) = \text{Var}\left(E(X | \vartheta_{\text{rv}})\right) + E\left(\text{Var}(X | \vartheta_{\text{rv}})\right).$$

Nun ist

$$(4) \quad \text{Var}(E(X|\vartheta_{\text{rv}})) = \text{Var}(n\vartheta_{\text{rv}}) = n^2 \text{Var}(\vartheta_{\text{rv}})$$

und weiter

$$\text{Var}(X|\vartheta_{\text{rv}}=\vartheta) = n\vartheta(1-\vartheta),$$

$$\text{Var}(X|\vartheta_{\text{rv}}) = n\vartheta_{\text{rv}}(1-\vartheta_{\text{rv}}) \quad (\text{bedingte Varianz als Zufallsgröße}),$$

$$(5) \quad \begin{aligned} E(\text{Var}(X|\vartheta_{\text{rv}})) &= nE[\vartheta_{\text{rv}}(1-\vartheta_{\text{rv}})] = n[E(\vartheta_{\text{rv}}) - E(\vartheta_{\text{rv}}^2)] = \\ &= nE(\vartheta_{\text{rv}})(1 - E(\vartheta_{\text{rv}})) - n\text{Var}(\vartheta_{\text{rv}}^2). \end{aligned}$$

Somit erhalten wir aus (3), (4) und (5)

$$\begin{aligned} \text{Var}(X) &= n^2 \text{Var}(\vartheta_{\text{rv}}) + nE(\vartheta_{\text{rv}})(1 - E(\vartheta_{\text{rv}})) - n\text{Var}(\vartheta_{\text{rv}}^2) = \\ &= (n^2 - n)\text{Var}(\vartheta_{\text{rv}}) + nE(\vartheta_{\text{rv}})(1 - E(\vartheta_{\text{rv}})) \\ &= n(n-1) \frac{ab}{(a+b)^2(a+b+1)} + n \frac{ab}{(a+b)^2} \\ &= \frac{nab}{(a+b)^2(a+b+1)} [(n-1) + (a+b+1)] \\ &= n(a+b+n) \frac{ab}{(a+b)^2(a+b+1)} = n(a+b+n)\text{Var}(\vartheta_{\text{rv}}). \end{aligned}$$

Im Spezialfall  $a = b = 1$  erhalten wir:

$$\begin{aligned} P\{X = k\} &= \binom{n}{k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)} \\ &= \frac{n!}{k!(n-k)!} \cdot \underbrace{\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}}_{=1} \cdot \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{n!}{(n+1)!} = \frac{1}{n+1} \end{aligned}$$

d.h.  $X$  besitzt eine diskrete Gleichverteilung auf  $\mathfrak{X} = \{0, 1, \dots, n\}$ . Aufgrund unserer Formeln ist

$$E(X) = n \frac{a}{a+b} = \frac{n}{2}$$

$$\text{Var}(X) = n(a+b+n) \frac{ab}{(a+b)^2(a+b+1)} = n(n+2) \frac{1}{4 \times 3} = \frac{1}{12} n(n+2)$$

und diese Ergebnisse stehen im Einklang mit der Gleichverteilung.

#### 1.4 Optimaler Bayes-Schätzer bei quadratischer Verlustfunktion

Es sei

$$\vartheta_{\text{rv}} \sim \text{Be}(a, b), \quad a, b \text{ bekannt};$$

$$X \sim \text{Bi}(n, \vartheta), \quad n \text{ bekannt, } \vartheta \text{ Realisation von } \vartheta_{\text{rv}};$$

$$\mathfrak{X} = \{0, 1, \dots, n\} \text{ Stichprobenraum zu } X;$$

$$\hat{\vartheta}: \mathfrak{X} \rightarrow [0, 1] \text{ Schätzfunktion für } \vartheta.$$

*Optimalitätskriterium:*

$$\hat{\vartheta} = \hat{\vartheta}(X) \text{ so wählen, daß } E\left(\hat{\vartheta}(X) - \vartheta_{\text{rv}}\right)^2 = \min$$

(globaler Erwartungswert bezüglich  $\vartheta_{\text{rv}}$  und  $X$ ).

*Lösung:*

Die optimale Schätzfunktion ist gegeben durch

$$\hat{\vartheta}(X) = E(\vartheta_{\text{rv}} | X) = \frac{a + X}{a + b + n};$$

sie ist also identisch zum Erwartungswert der a-posteriori-Verteilung von  $\vartheta_{\text{rv}}$ .

*Beweis:*

Es sei  $Y = \left(\hat{\vartheta}(X) - \vartheta_{\text{rv}}\right)^2$ . Aufgrund des Satzes über den bedingten Erwartungswert gilt

$$E(Y) = E\left(E(Y|X)\right). \text{ Nun ist}$$

$$\begin{aligned} E\left(E(Y|X)\right) &= \sum_{k=0}^n E(Y|X=k) P\{X=k\} = \sum_{k=0}^n E\left(\left(\hat{\vartheta}(X) - \vartheta_{\text{rv}}\right)^2 \middle| X=k\right) P\{X=k\} \\ (6) \quad &= \sum_{k=0}^n E\left(\left(\hat{\vartheta}(k) - \vartheta_{\text{rv}}\right)^2 \middle| X=k\right) P\{X=k\}, \end{aligned}$$

und für jedes  $k \in \mathfrak{X}$  gilt

$$E\left(\left(\hat{\vartheta}(k) - \vartheta_{\text{rv}}\right)^2 \middle| X=k\right) = \min$$

falls

$$(7) \quad \hat{\vartheta}(k) = E(\vartheta_{\text{rv}} | X=k) = \frac{a + k}{a + b + n},$$

denn für eine beliebige Zufallsgröße  $V$  gilt:

$$E(V - a)^2 = \text{Var}(V) + (\mu_V - a)^2 = \min \text{ falls } a = \mu_V = E(V)$$

(Minimumseigenschaft des Erwartungswerts). Da nun jeder einzelne Summand in der Summe (6) minimal wird für  $\hat{\vartheta}(k)$  gemäß (7), so wird auch die Gesamtsumme minimal, und damit ist die Behauptung bewiesen.

Aufgrund des Satzes über den bedingten Erwartungswert gilt

$$E\left(\hat{\vartheta}(X)\right) = E\left(E(\vartheta_{\text{rv}} | X)\right) = E(\vartheta_{\text{rv}}),$$

und in diesem Sinne kann die Schätzfunktion  $\hat{\vartheta}(X)$  als erwartungstreu bezeichnet werden.

Im Spezialfall  $a = b = 1$  gilt:

$$\hat{\vartheta}(X) = E(\vartheta_{\text{rv}} | X) = \frac{X + 1}{n + 2}.$$

## 2. Poissonverteilung mit Gamma-Verteilung als a-priori-Verteilung

### 2.1 Ausgangssituation

Es sei

$$X \sim \text{Po}(\lambda), \lambda \text{ unbekannt.}$$

Der unbekannte Parameter  $\lambda$  sei eine Realisation einer Zufallsgröße  $\lambda_{\text{rv}}$  (rv = random variable) mit einer Gammaverteilung:

$$\lambda_{\text{rv}} \sim \text{Ga}(a), a > 0, a \text{ bekannt};$$

$$\text{Dichte von } \lambda_{\text{rv}}: f(\lambda) = \frac{1}{\Gamma(a)} \lambda^{a-1} e^{-\lambda}, \lambda > 0;$$

$$E(\lambda_{\text{rv}}) = a,$$

$$\text{Var}(\lambda_{\text{rv}}) = a,$$

$$\text{Mo}(\lambda_{\text{rv}}) = a - 1 \text{ falls } a > 1.$$

Falls  $a = 1$ :  $f(\lambda) = e^{-\lambda}$  (einseitige Exponentialverteilung).

*Zweistufiges Experiment:*

1. Stufe:  $\lambda_{\text{rv}}$  realisieren gemäß  $\text{Ga}(a) \rightarrow \lambda > 0$ ;
2. Stufe:  $X$  realisieren gemäß  $\text{Po}(\lambda) \rightarrow k \in \{0, 1, \dots\}$ .

*Bayes'sche Statistik:*

$a$  bekannter Parameter (sogenannter Hyperparameter);

$\lambda$  unbekannte Realisation von  $\lambda_{\text{rv}}$ ;

$k$  vorliegende Beobachtung von  $X$ .

Ziel:  $\lambda$  schätzen; Vertrauensgrenzen für  $\lambda$  bestimmen.

*A-priori-Verteilung* von  $\lambda_{\text{rv}}$ :  $\text{Ga}(a)$ ;

*a-posteriori-Verteilung* von  $\lambda_{\text{rv}}$ : Bedingte Verteilung von  $\lambda_{\text{rv}}$  gegeben  $X = k$ .

### 2.2 A-posteriori-Verteilung von $\lambda_{\text{rv}}$

Wir betrachten also die folgenden Verteilungen:

$$\lambda_{\text{rv}} \sim \text{Ga}(a) \quad \text{stetige Verteilung;}$$

$$X \sim \text{Po}(\lambda) \quad \text{diskrete Verteilung;}$$

$$\lambda_{\text{rv}} | X=k \quad \text{stetige Verteilung.}$$

Die Dichte von  $\lambda_{\text{rv}} | X=k$  ist gegeben durch

$$\begin{aligned} (8) \quad f_k(\lambda) &= \frac{f(\lambda) p_\lambda(k)}{P\{X=k\}} = \frac{1}{P\{X=k\}} \cdot \frac{1}{\Gamma(a)} \lambda^{a-1} e^{-\lambda} \times \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \frac{1}{P\{X=k\}} \cdot \frac{1}{k! \Gamma(a)} \lambda^{a+k-1} e^{-2\lambda}. \end{aligned}$$

Nun ist

$$\begin{aligned}
 P\{X = k\} &= P\{\lambda_{\text{rv}} = \text{beliebiger Wert}, X = k\} = \int_0^{\infty} f(\lambda) p_{\lambda}(k) d\lambda \\
 (9) \quad &= \frac{1}{k! \Gamma(a)} \underbrace{\int_0^{\infty} \lambda^{a+k-1} e^{-2\lambda} d\lambda}_{= 2^{-(a+k)} \Gamma(a+k)} = \frac{\Gamma(a+k)}{2^{a+k} k! \Gamma(a)}
 \end{aligned}$$

und somit ergibt sich aus (8) und (9)

$$(10) \quad f_k(\lambda) = \frac{2}{\Gamma(a+k)} (2\lambda)^{a+k-1} e^{-2\lambda};$$

wenn  $Y$  eine Gamma-Verteilung  $\text{Ga}(a+k)$  besitzt, dann besitzt die Zufallsgröße  $\frac{1}{2}Y$  die Dichte  $f_k(\lambda)$  gemäß (10). Wir haben also das Ergebnis

$$\begin{aligned}
 \lambda_{\text{rv}} | X=k &\sim \frac{1}{2}Y, \text{ wobei } Y \sim \text{Ga}(a+k); \\
 E(\lambda_{\text{rv}} | X=k) &= \frac{1}{2}E(Y) = \frac{1}{2}(a+k); \\
 \text{Var}(\lambda_{\text{rv}} | X=k) &= \frac{1}{4}\text{Var}(Y) = \frac{1}{4}(a+k); \\
 \text{Mo}(\lambda_{\text{rv}} | X=k) &= \frac{1}{2}\text{Mo}(Y) = \frac{1}{2}(a+k-1).
 \end{aligned}$$

*A-posteriori-Modus-Schätzung* (=Maximum-Likelihood-Schätzung in der Bayes-Theorie):

$$\hat{\lambda} = \text{Mo}(\lambda_{\text{rv}} | X=k) = \frac{1}{2}(a+k-1)$$

Vertrauensgrenzen für  $\lambda$ :

$$\left. \begin{array}{l} \lambda_{\ell} = \text{linkes} \\ \lambda_r = \text{rechtes} \end{array} \right\} \alpha\text{-Quantil der a-posteriori-Verteilung von } \lambda_{\text{rv}} | X=k, \text{ d.h.}$$

$$\lambda_{\ell} = \frac{1}{2}\Gamma_{\alpha}(a+k);$$

$$\lambda_r = \frac{1}{2}\Gamma_{1-\alpha}(a+k).$$

$$\text{Es gilt dann: } P\{\lambda_{\ell} < \lambda_{\text{rv}} | X=k < \lambda_r\} = 1 - 2\alpha.$$

Im Spezialfall  $a=1$  gilt:

$$\lambda_{\text{rv}} \text{ besitzt die Dichte } f(\lambda) = e^{-\lambda};$$

$$\lambda_{\text{rv}} | X=k \sim \frac{1}{2}Y, \text{ wobei } Y \sim \text{Ga}(k+1);$$

$$\hat{\lambda} = \text{Mo}(\lambda_{\text{rv}} | X=k) = \frac{1}{2}k;$$

$$\lambda_{\ell} = \frac{1}{2}\Gamma_{\alpha}(k+1);$$

$$\lambda_r = \frac{1}{2}\Gamma_{1-\alpha}(k+1).$$

Klassische Vertrauensgrenzen für  $\lambda$  (mit Konfidenzinterpretation):

$$\lambda_{\ell} = \Gamma_{\alpha}(k) \quad \text{falls } k > 0 \quad (\lambda_{\ell} = 0 \text{ falls } k = 0);$$

$$\lambda_r = \Gamma_{1-\alpha}(k+1).$$

### 2.3 Nicht-informative a-priori-Verteilung

Da eine nicht-informative a-priori-Verteilung alle Parameter  $\lambda \in (0, \infty)$  gleich behandeln sollte, und da es keine (auf 1 normierte) Gleichverteilung auf dem Intervall  $(0, \infty)$  gibt, so liegt es nahe das Lebesgue-Maß, welches jedem Intervall  $(a, b)$  die Länge  $b - a$  als Maß zuordnet, als nicht-informative a-priori-Verteilung zu betrachten (diffuse a-priori-Verteilung). Wir betrachten also die folgenden "Verteilungen":

$\lambda_{\text{rv}}$  = "Zufallsgröße" mit Gleichverteilung (Lebesgue-Maß) auf  $(0, \infty)$ ;

$X \sim \text{Po}(\lambda)$  diskrete Verteilung;

$\lambda_{\text{rv}} | X=k$  stetige Verteilung.

Die "Dichte" der a-priori-Verteilung von  $\lambda_{\text{rv}}$  ist gegeben durch  $f(\lambda) = 1$  für  $\lambda \in (0, \infty)$ . Die "Dichte" von  $\lambda_{\text{rv}} | X=k$  ist dann gegeben durch

$$(11) \quad f_k(\lambda) = \frac{f(\lambda) p_\lambda(k)}{P\{X=k\}} = \frac{p_\lambda(k)}{P\{X=k\}} = \frac{1}{P\{X=k\}} \cdot \frac{e^{-\lambda} \lambda^k}{k!}.$$

Nun ist die "Marginalverteilung" von  $X$  gegeben durch

$$(12) \quad \begin{aligned} P\{X=k\} &= P\{\lambda_{\text{rv}} = \text{beliebiger Wert}, X=k\} = \int_0^\infty f(\lambda) p_\lambda(k) d\lambda \\ &= \int_0^\infty p_\lambda(k) d\lambda = \frac{1}{k!} \underbrace{\int_0^\infty \lambda^k e^{-\lambda} d\lambda}_0 = \frac{\Gamma(k+1)}{k!} = 1. \end{aligned}$$

Auch die "Marginalverteilung" von  $X$  ist nun eine "Gleichverteilung" auf  $\mathfrak{X} = \{0, 1, \dots\}$ . Gemäß (11) ist die a-posteriori-Verteilung von  $\lambda_{\text{rv}} | X=k$  gegeben durch

$$f_k(\lambda) = \frac{1}{P\{X=k\}} \cdot \frac{e^{-\lambda} \lambda^k}{k!} = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Da gemäß (12)

$$\int_0^\infty f_k(\lambda) d\lambda = \frac{1}{k!} \int_0^\infty e^{-\lambda} \lambda^k d\lambda = 1,$$

so ist  $f_k(\lambda)$  eine korrekte Wahrscheinlichkeitsdichte für jedes  $k \in \{0, 1, \dots\}$ . Wir haben also das Ergebnis

$$\lambda_{\text{rv}} | X=k \sim \text{Ga}(k+1) \text{ mit der Dichte } f_k(\lambda) = \frac{e^{-\lambda} \lambda^k}{k!};$$

$$E(\lambda_{\text{rv}} | X=k) = \int_0^\infty \lambda f(\lambda) d\lambda = k+1;$$

$$\text{Var}(\lambda_{\text{rv}} | X=k) = \int_0^\infty \lambda^2 f(\lambda) d\lambda - (k+1)^2 = k+1;$$

$$\text{Mo}(\lambda_{\text{rv}} | X=k) = k.$$

*A-posteriori-Modus-Schätzung* (=Maximum-Likelihood-Schätzung in der Bayes-Theorie):

$$\hat{\lambda} = \text{Mo}(\lambda_{\text{rv}} | X=k) = k \quad (\text{wie im klassischen Fall}).$$

Vertrauensgrenzen für  $\lambda$  :

$$\left. \begin{array}{l} \lambda_\ell = \text{linkes} \\ \lambda_r = \text{rechtes} \end{array} \right\} \alpha\text{-Quantil der a-posteriori-Verteilung } \text{Ga}(k+1), \text{ d.h.}$$

$$\lambda_\ell = \Gamma_\alpha(k+1);$$

$$\lambda_r = \Gamma_{1-\alpha}(k+1).$$

$$\text{Es gilt dann: } P\{\lambda_\ell < \lambda_{\text{rv}} | X=k < \lambda_r\} = 1 - 2\alpha.$$

Die klassischen Vertrauensgrenzen (mit Konfidenzinterpretation) sind

$$\lambda_\ell = \Gamma_\alpha(k);$$

$$\lambda_r = \Gamma_{1-\alpha}(k+1).$$

*Bemerkung:*

Die Bayes-Schätzung  $\hat{\lambda} = k$  stimmt also hier mit der klassischen Maximum-Likelihood-Schätzung überein, und die linke Vertrauensgrenze nach Bayes ist stets etwas größer (schärfer) als im klassischen Fall. Die klassische Konfidenzinterpretation ist jedoch korrekt für beliebige Werte des unbekanntes Parameters  $\lambda$  und somit auch bei Vorliegen einer *beliebigen* a-priori-Verteilung für  $\lambda$ .

## 2.4 Randverteilung von $X$ ; Rechnen mit bedingten Erwartungswerten

Wir betrachten das zweistufige Bayes-Experiment (wie oben unter 2.1) und wollen die Randverteilung (Marginalverteilung) von  $X$  untersuchen. Gemäß (9) gilt:

$$P\{X = k\} = P\{\lambda_{\text{rv}} = \text{beliebiger Wert}, X = k\} = \int_0^\infty f(\lambda) p_\lambda(k) d\lambda = \frac{\Gamma(a+k)}{2^{a+k} k! \Gamma(a)}$$

Nun wollen wir den Erwartungswert und die Varianz von  $X$  berechnen mit Hilfe der Sätze über den bedingten Erwartungswert und die bedingte Varianz. Der Satz vom bedingten Erwartungswert lautet:

$$E(X) = E(E(X | \lambda_{\text{rv}})).$$

Nun ist

$$E(X | \lambda_{\text{rv}} = \lambda) = \lambda,$$

$$E(X | \lambda_{\text{rv}}) = \lambda_{\text{rv}} \text{ (bedingter Erwartungswert als Zufallsgröße),}$$

und somit

$$E(X) = E(E(X | \lambda_{\text{rv}})) = E(\lambda_{\text{rv}}) = a.$$

Der Satz von der bedingten Varianz lautet:

$$(13) \text{Var}(X) = \text{Var}(E(X | \lambda_{\text{rv}})) + E(\text{Var}(X | \lambda_{\text{rv}})).$$

Nun ist

$$(14) \text{Var}(E(X | \lambda_{\text{rv}})) = \text{Var}(\lambda_{\text{rv}}) = a$$

und weiter

$$\begin{aligned} \text{Var}(X|\lambda_{\text{rv}}=\lambda) &= \lambda, \\ (15) \quad \text{Var}(X|\lambda_{\text{rv}}) &= \lambda_{\text{rv}} \quad (\text{bedingte Varianz als Zufallsgröße}), \\ E(\text{Var}(X|\lambda_{\text{rv}})) &= E(\lambda_{\text{rv}}) = a. \end{aligned}$$

Somit erhalten wir aus (13), (14) und (15)

$$\text{Var}(X) = a + a = 2a$$

Im Spezialfall  $a = 1$  erhalten wir:

$$P\{X = k\} = \frac{\Gamma(a+k)}{2^{a+k} k! \Gamma(a)} = \frac{\Gamma(k+1)}{2^{k+1} k! \Gamma(1)} = \frac{1}{2^{k+1}}, \quad k = 0, 1, \dots$$

d.h.  $X$  besitzt eine geometrische Verteilung auf  $\mathfrak{X} = \{0, 1, \dots\}$ . Aufgrund unserer Formeln ist

$$\begin{aligned} E(X) &= a = 1 \\ \text{Var}(X) &= 2a = 2 \end{aligned}$$

und diese Ergebnisse stehen im Einklang mit der geometrischen Verteilung.

## 2.5 Optimaler Bayes-Schätzer bei quadratischer Verlustfunktion

Es sei

$$\begin{aligned} \lambda_{\text{rv}} &\sim \text{Ga}(a), \quad a \text{ bekannt}; \\ X &\sim \text{Po}(\lambda), \quad \lambda \text{ Realisation von } \lambda_{\text{rv}}; \\ \mathfrak{X} &= \{0, 1, \dots\} \quad \text{Stichprobenraum zu } X; \\ \hat{\lambda}: \mathfrak{X} &\rightarrow [0, \infty) \quad \text{Schätzfunktion für } \lambda. \end{aligned}$$

*Optimalitätskriterium:*

$$\begin{aligned} \hat{\lambda} = \hat{\lambda}(X) \text{ so wählen: } &E\left(\hat{\lambda}(X) - \lambda_{\text{rv}}\right)^2 = \min \\ &(\text{globaler Erwartungswert bezüglich } \lambda_{\text{rv}} \text{ und } X). \end{aligned}$$

*Lösung:*

Die optimale Schätzfunktion ist gegeben durch

$$\hat{\lambda}(X) = E(\lambda_{\text{rv}}|X) = \frac{1}{2}(a + X);$$

sie ist also identisch zum Erwartungswert der a-posteriori-Verteilung von  $\lambda_{\text{rv}}$ .

*Beweis:*

Es sei  $Y = \left(\hat{\lambda}(X) - \lambda_{\text{rv}}\right)^2$ . Aufgrund des Satzes über den bedingten Erwartungswert gilt  $E(Y) = E(E(Y|X))$ . Nun ist

$$\begin{aligned} (16) \quad E(E(Y|X)) &= \sum_{k=0}^{\infty} E(Y|X=k) P\{X=k\} = \sum_{k=0}^{\infty} E\left(\left(\hat{\lambda}(X) - \lambda_{\text{rv}}\right)^2 \middle| X=k\right) P\{X=k\} \\ &= \sum_{k=0}^{\infty} E\left(\left(\hat{\lambda}(k) - \lambda_{\text{rv}}\right)^2 \middle| X=k\right) P\{X=k\}, \end{aligned}$$

und für jedes  $k \in \mathfrak{X}$  gilt

$$E\left(\left(\hat{\lambda}(k) - \lambda_{\text{TV}}\right)^2 \middle| X=k\right) = \min$$

falls

$$(17) \quad \hat{\lambda}(k) = E(\lambda_{\text{TV}} | X=k) = \frac{1}{2}(a+k),$$

(Minimumseigenschaft des Erwartungswerts). Da nun jeder einzelne Summand in der Summe (16) minimal wird für  $\hat{\lambda}(k)$  gemäß (17), so wird auch die Gesamtsumme minimal, und damit ist die Behauptung bewiesen.

Aufgrund des Satzes über den bedingten Erwartungswert gilt

$$E(\hat{\lambda}(X)) = E(E(\lambda_{\text{TV}} | X)) = E(\lambda_{\text{TV}}),$$

und in diesem Sinne kann die Schätzfunktion  $\hat{\lambda}(X)$  als erwartungstreu bezeichnet werden.

### 3. Multinomialverteilung mit Dirichlet-Verteilung als a-priori-Verteilung

#### 3.1 Dirichlet-Verteilung

Die Dirichlet-Verteilung ist eine Verallgemeinerung der Beta-Verteilung. Der  $r$ -dimensionale Zufallsvektor  $X = (X_1, \dots, X_r)$  besitzt eine Dirichlet-Verteilung mit den Parametern  $a_1, \dots, a_{r+1}$  (alle  $a_i > 0$ ), falls die Dichte gegeben ist durch

$$f(x_1, \dots, x_r) = \frac{\Gamma(a_\bullet)}{\Gamma(a_1) \cdots \Gamma(a_{r+1})} x_1^{a_1-1} \cdots x_{r+1}^{a_{r+1}-1},$$

wobei

$$a_\bullet = a_1 + \cdots + a_{r+1};$$

$$0 \leq x_i \leq 1 \quad \text{für } i = 1, \dots, r;$$

$$\sum_{i=1}^r x_i \leq 1;$$

$$x_{r+1} = 1 - \sum_{i=1}^r x_i.$$

Wir schreiben kurz:  $X = (X_1, \dots, X_r) \sim \text{Di}(a_1, \dots, a_{r+1})$ . Für  $r=1$  ist die Dirichlet-Verteilung  $\text{Di}(a, b)$  identisch zur Beta-Verteilung  $\text{Be}(a, b)$ . Der Stichprobenraum  $\mathfrak{X}$  (= Wertebereich des Zufallsvektors  $X$ ) ist gegeben durch

$$\mathfrak{X} = \{(x_1, \dots, x_r) \mid 0 \leq x_i \leq 1 \text{ für } i = 1, \dots, r; x_1 + \cdots + x_r \leq 1\}.$$

Die Menge  $\mathfrak{X} (\subset \mathbb{R}^r)$  heißt ein Simplex; für  $r=2$  hat  $\mathfrak{X}$  die Form eines Dreiecks, für  $r=3$  die Form einer dreiseitigen Pyramide. Weiter gilt für jede Komponente  $X_i$ :

$$(18) \quad X_i \sim \text{Be}(\alpha, \beta), \quad \text{wobei } \alpha = a_i, \beta = a_\bullet - a_i.$$

Daraus folgt:

$$E(X_i) = \mu_i = \frac{\alpha}{\alpha + \beta} = \frac{a_i}{a_\bullet}; \quad \text{Mo}(X_i) = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{a_i - 1}{a_\bullet - 2} \quad \text{falls } a_i > 1, a_\bullet > 2;$$

$$\text{Var}(X_i) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{a_i(a_\bullet - a_i)}{a_\bullet^2(a_\bullet + 1)} = \frac{\mu_i(1 - \mu_i)}{a_\bullet + 1}.$$

Begründung für (18):

1° Es sei  $(X, Y) \sim \text{Di}(a, b, c)$ . Dann ist die Marginalverteilung von  $X$  gegeben durch

$$f_X(x) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} x^{a-1} \int_0^{1-x} y^{b-1} (1-x-y)^{c-1} dy.$$

Mit der Substitution  $t = \frac{y}{1-x}$  erhalten wir

$$\begin{aligned} f_X(x) &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} x^{a-1} (1-x)^{b+c-1} \underbrace{\int_0^1 t^{b-1} (1-t)^{c-1} dt}_{\frac{\Gamma(b)\Gamma(c)}{\Gamma(b+c)}} \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b+c)} x^{a-1} (1-x)^{b+c-1}, \end{aligned}$$

d.h.  $X$  besitzt eine Beta-Verteilung  $\text{Be}(a, b+c)$ .

2° Es sei  $(X, Y, Z) \sim \text{Di}(a, b, c, d)$ . Dann ist die Verteilung von  $(X, Y)$  gegeben durch

$$f_{XY}(x, y) = \frac{\Gamma(a+b+c+d)}{\Gamma(a)\Gamma(b)\Gamma(c)\Gamma(d)} x^{a-1} y^{b-1} \int_0^{1-x-y} z^{c-1} (1-x-y-z)^{d-1} dz.$$

Mit der Substitution  $t = \frac{z}{1-x-y}$  erhalten wir

$$\begin{aligned} f_{XY}(x, y) &= \frac{\Gamma(a+b+c+d)}{\Gamma(a)\Gamma(b)\Gamma(c)\Gamma(d)} x^{a-1} y^{b-1} (1-x-y)^{c+d-1} \underbrace{\int_0^1 t^{c-1} (1-t)^{d-1} dt}_{\frac{\Gamma(c)\Gamma(d)}{\Gamma(c+d)}} \\ &= \frac{\Gamma(a+b+c+d)}{\Gamma(a)\Gamma(b)\Gamma(c+d)} x^{a-1} y^{b-1} (1-x-y)^{c+d-1}, \end{aligned}$$

d.h.  $(X, Y)$  besitzt eine Dirichlet-Verteilung  $\text{Di}(a, b, c+d)$ .

3° Der Beweisschritt 2° gilt analog auch für  $r$ -dimensionale Dirichlet-Verteilungen mit  $r > 3$ , und damit ist unsere Behauptung bewiesen.

### 3.2 Bayes-Modell

Es sei

$$Y = (Y_1, \dots, Y_{r+1}) \sim \text{Mu}(n, \vartheta_1, \dots, \vartheta_{r+1})$$

mit bekanntem  $n$  und unbekanntem Parametern  $\vartheta_1, \dots, \vartheta_{r+1}$ . Es gilt  $0 \leq \vartheta_i \leq 1$  für  $i = 1, \dots, r+1$  und  $\vartheta_1 + \dots + \vartheta_{r+1} = 1$ . Da  $Y_1 + \dots + Y_{r+1} = n$ , so ist  $Y$  eigentlich nur ein  $r$ -dimensionaler Zufallsvektor. Nun seien die unbekanntem Parameter  $\vartheta_1, \dots, \vartheta_{r+1}$  Realisationen eines Zufallsvektors mit einer Dirichlet-Verteilung:

$$X = (X_1, \dots, X_r) \sim \text{Di}(a_1, \dots, a_{r+1}),$$

wobei die sogenannten Hyperparameter  $a_1, \dots, a_{r+1}$  als bekannt vorausgesetzt werden.

*Zweistufiges Experiment:*

1. Stufe:  $X = (X_1, \dots, X_r)$  realisieren gemäß  $\text{Di}(a_1, \dots, a_{r+1})$   
 $\rightarrow x = (\vartheta_1, \dots, \vartheta_r)$  mit  $0 \leq \vartheta_i \leq 1, \vartheta_1 + \dots + \vartheta_r \leq 1$ ;  
wir setzen  $\vartheta_{r+1} = 1 - (\vartheta_1 + \dots + \vartheta_r)$ .
2. Stufe:  $Y = (Y_1, \dots, Y_{r+1})$  realisieren gemäß  $\text{Mu}(n, \vartheta_1, \dots, \vartheta_{r+1})$   
 $\rightarrow y = (n_1, \dots, n_{r+1})$  mit  $n_i \in \{0, 1, \dots, n\}, n_1 + \dots + n_{r+1} = n$ .

*Bayes'sche Statistik:*

$n, a_1, \dots, a_{r+1}$  bekannte Parameter ( $a_1, \dots, a_{r+1}$  sogenannte Hyperparameter);

$x = (\vartheta_1, \dots, \vartheta_r)$  unbekannte Realisation von  $X = (X_1, \dots, X_r)$ ;

$y = (n_1, \dots, n_{r+1})$  vorliegende Beobachtung von  $Y = (Y_1, \dots, Y_{r+1})$ .

Ziel:  $\vartheta_i$  schätzen, Vertrauensgrenzen für  $\vartheta_i$  bestimmen ( $i = 1, \dots, r+1$ ).

*A-priori-Verteilung* von  $X = (X_1, \dots, X_r)$ :  $\text{Di}(a_1, \dots, a_{r+1})$ ;

*a-posteriori-Verteilung* von  $X = (X_1, \dots, X_r)$ : Bedingte Verteilung von  $X$  gegeben  $Y = y$ .

*Bestimmung der a-posteriori-Verteilung von X:*

Die Dichte von  $X|Y=y$  ist gegeben durch

$$\begin{aligned}
 f_y(x) &= \frac{f(x)p_\vartheta(y)}{P(Y=y)} \\
 (19) \quad &= \frac{1}{P(Y=y)} \cdot \frac{\Gamma(a_\bullet)}{\Gamma(a_1) \cdots \Gamma(a_{r+1})} \vartheta_1^{a_1-1} \cdots \vartheta_{r+1}^{a_{r+1}-1} \frac{n!}{n_1! \cdots n_{r+1}!} \vartheta_1^{n_1} \cdots \vartheta_{r+1}^{n_{r+1}} \\
 &= \frac{1}{P(Y=y)} \cdot \frac{\Gamma(a_\bullet)}{\Gamma(a_1) \cdots \Gamma(a_{r+1})} \cdot \frac{n!}{n_1! \cdots n_{r+1}!} \vartheta_1^{a_1+n_1-1} \cdots \vartheta_{r+1}^{a_{r+1}+n_{r+1}-1}
 \end{aligned}$$

Als Marginalverteilung von  $Y$  erhalten wir

$$\begin{aligned}
 P\{Y=y\} &= P\{X = \text{beliebiger Wert}, Y=y\} = \int_{\mathfrak{X}} f(x)p_\vartheta(y)dx \\
 (20) \quad &= \frac{\Gamma(a_\bullet)}{\Gamma(a_1) \cdots \Gamma(a_{r+1})} \cdot \frac{n!}{n_1! \cdots n_{r+1}!} \int_{\mathfrak{X}} \vartheta_1^{a_1+n_1-1} \cdots \vartheta_{r+1}^{a_{r+1}+n_{r+1}-1} d\vartheta_1 \cdots d\vartheta_r \\
 &= \frac{\Gamma(a_\bullet)}{\Gamma(a_1) \cdots \Gamma(a_{r+1})} \cdot \frac{n!}{n_1! \cdots n_{r+1}!} \cdot \frac{\Gamma(a_1+n_1) \cdots \Gamma(a_{r+1}+n_{r+1})}{\Gamma(a_\bullet+n)}
 \end{aligned}$$

Aus (19) und (20) erhalten wir

$$f_y(x) = \frac{\Gamma(a_\bullet+n)}{\Gamma(a_1+n_1) \cdots \Gamma(a_{r+1}+n_{r+1})} \vartheta_1^{a_1+n_1-1} \cdots \vartheta_{r+1}^{a_{r+1}+n_{r+1}-1},$$

und dies ist die Dichte der Dirichlet-Verteilung  $\text{Di}(a_1+n_1, \dots, a_{r+1}+n_{r+1})$ .

Wir haben also das Ergebnis

$$X|Y=y \sim \text{Di}(a_1+n_1, \dots, a_{r+1}+n_{r+1}),$$

$$X_i|Y=y \sim \text{Be}(\alpha, \beta) \text{ mit } \alpha = a_i + n_i, \beta = a_\bullet + n - a_i - n_i,$$

und damit

$$E(X_i|Y=y) = \mu_i = \frac{a_i + n_i}{a_\bullet + n} \quad (\approx p_i = \frac{n_i}{n}, \text{ falls } n \text{ und } n_i \text{ groß});$$

$$\text{Var}(X_i|Y=y) = \frac{\mu_i(1-\mu_i)}{a_\bullet + n + 1} \quad (\approx \frac{p_i(1-p_i)}{n}, \text{ falls } n \text{ und } n_i \text{ groß});$$

$$\text{Mo}(X_i|Y=y) = \frac{a_i + n_i - 1}{a_\bullet + n - 2} \quad (\approx p_i = \frac{n_i}{n}, \text{ falls } n \text{ und } n_i \text{ groß}).$$

Punktschätzungen und Vertrauensgrenzen für die einzelnen Parameter  $\vartheta_i$ ,  $i = 1, \dots, r+1$  können nun wie im ersten Beispiel (Binomialverteilung und Beta-Verteilung) bestimmt werden. Die a-posteriori-Modus Schätzung für  $\vartheta_i$  ist also

$$\hat{\vartheta}_i = \text{Mo}(X_i|Y=y) = \frac{a_i + n_i - 1}{a_\bullet + n - 2}.$$

*Nicht-informative a-priori-Verteilung:*

Als nicht-informative a-priori-Verteilung kommt die Gleichverteilung auf dem Simplex  $\mathfrak{X}$  in Frage. Dies ist die Dirichlet-Verteilung  $\text{Di}(a_1, \dots, a_{r+1})$  mit  $a_1 = a_2 = \dots = a_{r+1} = 1$ . Die Dichte dieser Verteilung ist  $f(x) = \Gamma(a_\bullet) = \Gamma(r+1) = r!$  für jedes  $x \in \mathfrak{X}$ . Damit ist das Volumen des Simplex  $\mathfrak{X}$  gegeben durch  $1/r!$ . Die a-posteriori-Modus-Schätzung für  $\vartheta_i$  ist jetzt

$$\hat{\vartheta}_i = \text{Mo}(X_i|Y=y) = \frac{a_i + n_i - 1}{a_\bullet + n - 2} = \frac{n_i}{n + r - 1}.$$

Nur im Falle  $r=1$  stimmt diese Schätzung mit der klassischen Maximum-Likelihood-Schätzung überein.

*Schlußbemerkung:*

Wir haben in diesem Kapitel drei Paare von konjugierten (zusammengehörigen) Verteilungen betrachtet:

*Beispiel 1:* Binomial- und Beta-Verteilung;

*Beispiel 2:* Poisson- und Gamma-Verteilung;

*Beispiel 3:* Multinomial- und Dirichlet-Verteilung.

Bei diesen Verteilungspaaren waren wir in der Lage, bei gegebener a-priori-Verteilung die a-posteriori-Verteilung exakt zu berechnen, und diese Verteilung gehörte wieder zur gleichen Familie wie die a-priori-Verteilung. Im Rahmen der Bayes'schen Statistik werden Schlüsse bezüglich des unbekanntes Parameters gezogen aufgrund der a-posteriori-Verteilung. Bayes-Modelle erfordern jedoch nach meiner Meinung (willkürliche) Annahmen über eine a-priori-Verteilung; auch wenn eine *nicht-informative* (oder *diffuse*) a-priori-Verteilung vorausgesetzt wird, so ist dies eine einschränkende Annahme. Bayes-Modelle sind kein Ersatz für die klassischen Verfahren (Maximum-Likelihood-Methode, Vertrauensgrenzen mit Konfidenzinterpretation).

## 19. Stochastische Ordnung, Test von Wilcoxon

Wir haben bereits verschiedene sogenannte *verteilungsfreie* Verfahren kennengelernt:

Median-Verfahren (Vertrauensgrenzen, Ein- und Zwei-Stichprobenproblem);

Tests von Kolmogoroff und Smirnow (Ein- und Zwei-Stichprobenproblem).

Hier wollen wir ein weiteres verteilungsfreies Verfahren zum Zwei-Stichprobenproblem vorstellen, den *Test von Wilcoxon*.

### 1. Stochastische Ordnung

*Stochastische Ordnung, Definition I:*

$X$  und  $Y$  seien zwei unabhängige (eindimensionale) Zufallsgrößen mit den Verteilungsfunktion  $F$  und  $G$ . Die Zufallsgröße  $X$  heißt *stochastisch größer* als  $Y$ , falls  $P\{X > Y\} > P\{X < Y\}$ . Wir schreiben dann:  $X \stackrel{I}{\succ} Y$ .

*Stochastische Ordnung, Definition II:*

$X$  und  $Y$  seien zwei (eindimensionale) Zufallsgrößen mit den Verteilungsfunktion  $F$  und  $G$ . Die Zufallsgröße  $X$  heißt *stochastisch größer* als  $Y$ , falls

$$F^c(x) \geq G^c(x) \quad \forall x \in \mathbb{R} \quad \text{und} \quad F \neq G.$$

Wir schreiben dann:  $X \stackrel{II}{\succ} Y$ .

*Folgerung:* Es gilt

$$\begin{aligned} X \stackrel{II}{\succ} Y &\Rightarrow P\{X > c\} \geq P\{Y > c\} \quad \forall c \in \mathbb{R} \\ &\Rightarrow P\{X \leq c\} \leq P\{Y \leq c\} \quad \forall c \in \mathbb{R} \\ &\Rightarrow F(c) \leq G(c) \quad \forall c \in \mathbb{R} \end{aligned}$$

Daraus ergibt sich die scheinbar unsinnige Folgerung, daß die Verteilungsfunktion von  $X$  *nie größer* ist als die Verteilungsfunktion von  $Y$ , wenn  $X$  stochastisch *größer* ist als  $Y$ .

*Satz: Es gilt*

$$(1) \quad X \stackrel{II}{\succ} Y \Rightarrow X \stackrel{I}{\succ} Y,$$

aber nicht umgekehrt. Jedes Paar  $(X, Y)$  von Zufallsgrößen, welches geordnet ist nach Definition II, ist auch geordnet (im gleichen Sinne) nach Definition I. Die Ordnung I ist also eine Verfeinerung der Ordnung II.

*Beweis:*

<sup>1°</sup> Es sei  $X$  eine beliebige (eindimensionale) Zufallsgröße mit der Verteilungsfunktion  $F$ ; die Verteilungsfunktion braucht nicht stetig zu sein, sie kann Sprungstellen besitzen. Wir schreiben

$$\begin{aligned} F(x-0) &= P\{X < x\}; \\ F(x+0) &= F(x) = P\{X \leq x\}. \end{aligned}$$

Nun ist  $H(x) = F^2(x)$  wiederum eine korrekte Verteilungsfunktion, und wenn  $x$  eine Sprungstelle von  $F$  und somit auch von  $H$  ist, so gilt

$$\begin{aligned} dH(x) &= H(x+0) - H(x-0) = F^2(x+0) - F^2(x-0) \\ &= [F(x+0) + F(x-0)] \cdot [F(x+0) - F(x-0)] = [F(x+0) + F(x-0)] dF(x). \end{aligned}$$

Daraus folgt

$$\int dH(x) = \int [F(x+0) + F(x-0)] dF(x) = 1.$$

2°  $X$  und  $Y$  seien zwei unabhängige Zufallsgrößen mit den Verteilungsfunktion  $F$  und  $G$ . Dann gilt

$$\begin{aligned} P\{X > Y\} &= \int G(x-0) dF(x); \\ P\{X < Y\} &= \int [1 - G(x+0)] dF(x) = 1 - \int G(x+0) dF(x); \\ P\{X = Y\} &= 1 - P\{X < Y\} - P\{X > Y\} = \int [G(x+0) - G(x-0)] dF(x). \end{aligned}$$

3°  $X$  und  $Y$  seien zwei unabhängige Zufallsgrößen mit den Verteilungsfunktion  $F$  und  $G$ , und es sei  $X \stackrel{\text{II}}{\succ} Y$ . Dann gilt:  $F(x) \leq G(x) \quad \forall x \in \mathbb{R}$  und  $F \not\equiv G$ ; ebenso gilt

$F(x-0) \leq G(x-0) \quad \forall x \in \mathbb{R}$ . Daraus folgt

$$\begin{aligned} (2) \quad P\{X > Y\} &= \int G(x-0) dF(x) \geq \int F(x-0) dF(x), \\ P\{X < Y\} &= 1 - \int G(x+0) dF(x) \leq 1 - \int F(x+0) dF(x), \end{aligned}$$

und wir erhalten

$$(3) \quad P\{X > Y\} - P\{X < Y\} \geq \underbrace{\int [F(x-0) + F(x+0)] dF(x)}_{=1 \text{ (gemäß 1°)}} - 1 = 0,$$

d.h.  $P\{X > Y\} \geq P\{X < Y\}$ . Wir haben noch zu zeigen, daß in der letzten Ungleichung das Gleichheitszeichen nicht möglich ist.

4° Es sei  $F(x) \leq G(x) \quad \forall x \in \mathbb{R}$  und

$$\int G(x) dF(x) = \int F(x) dF(x).$$

Wenn  $F$  diskret ist mit den Sprungstellen  $x_1, x_2, \dots$ , dann gilt

$$\int [G(x) - F(x)] dF(x) = \sum_i \underbrace{[G(x_i) - F(x_i)]}_{\geq 0} dF(x_i) = 0,$$

und daraus folgt  $G(x_i) = F(x_i)$  für  $i = 1, 2, \dots$ , d.h.  $F \equiv G$ . Wenn  $F$  eine stetige Dichte  $f$  besitzt, dann gilt

$$\int [G(x) - F(x)] dF(x) = \int \underbrace{[G(x) - F(x)]}_{\geq 0} f(x) dx = 0,$$

und daraus folgt  $G(x) = F(x) \quad \forall x$  mit  $f(x) > 0$ ; wegen der Monotonie der Verteilungsfunktionen folgt daraus  $G(x) = F(x) \quad \forall x \in \mathbb{R}$ . Da in den Beziehungen (2) und (3)  $F \not\equiv G$  ist, so müssen also die Ungleichungen in diesen Beziehungen echt sein, da das Gleichheitszeichen nicht möglich ist, und damit ist die Behauptung (1) bewiesen.

*Beispiel:*

Wir wollen durch ein Gegenbeispiel zeigen, dass die Umkehrung der Behauptung (1) nicht richtig ist.

Es sei  $X$  eine stetige Zufallsgröße mit der Dichte

$$f(x) = \begin{cases} \frac{1}{2} & \text{für } 0 < x \leq \frac{2}{5} \\ 2 & \text{für } \frac{2}{5} < x \leq \frac{4}{5} \\ 0 & \text{sonst,} \end{cases}$$

und  $Y$  besitze eine Gleichverteilung im Intervall  $[0,1]$ . Dann gilt

$$P\{X > Y\} = \int G(x)f(x) dx = \frac{1}{2} \int_0^{\frac{2}{5}} x dx + 2 \int_{\frac{2}{5}}^{\frac{4}{5}} x dx = \frac{13}{25} > \frac{1}{2},$$

d.h.  $X \succ^I Y$ . Es ist jedoch  $F(\frac{2}{5}) = \frac{1}{5} < G(\frac{2}{5}) = \frac{2}{5}$  und  $F(\frac{4}{5}) = 1 > G(\frac{4}{5}) = \frac{4}{5}$ , und somit gilt nicht  $F(x) \leq G(x) \quad \forall x \in \mathbb{R}$ , d.h.  $X \succ^I Y \not\Rightarrow X \succ^{II} Y$ .

*Resultat:*

Die stochastische Ordnung gemäß Definition I ist eine *vollständige* Ordnung; für jedes Paar  $(X, Y)$  von Zufallsgrößen gilt

$$\text{entweder } P\{X > Y\} > P\{X < Y\}, \text{ d.h. } X \succ^I Y,$$

$$\text{oder } P\{X > Y\} < P\{X < Y\}, \text{ d.h. } X \prec^I Y,$$

$$\text{oder } P\{X > Y\} = P\{X < Y\}, \text{ d.h. } X \stackrel{I}{=} Y.$$

Die stochastische Ordnung gemäß Definition II ist keine vollständige sondern nur eine *partielle* Ordnung; es gibt Paare  $(F, G)$  von Verteilungsfunktionen, welche nicht geordnet sind gemäß Definition II (siehe das obige Beispiel). Die Ordnung gemäß Definition II ist eine Vergrößerung der Ordnung gemäß Definition I, denn alle Paare von Zufallsgrößen  $X, Y$  (mit zugehörigen Verteilungsfunktionen  $F$  und  $G$ ), welche geordnet sind gemäß Definition II, sind auch im gleichen Sinne geordnet gemäß Definition I, aber es gibt Paare, welche nicht geordnet sind gemäß Definition II, wohl aber gemäß Definition I (siehe das obige Beispiel).

## 2. Test von Wilcoxon (-Mann-Whitney); Problemstellung

Stochastische Ordnung gemäß Definition I:

$X$  und  $Y$  seien unabhängige Zufallsgrößen mit den Verteilungsfunktionen  $F$  und  $G$ . Wir schreiben

$$F \stackrel{\text{st}}{\neq} G \Leftrightarrow X \stackrel{\text{I}}{\neq} Y.$$

Nun seien

$$\left. \begin{array}{l} X_1, \dots, X_{n_1} \\ Y_1, \dots, Y_{n_2} \end{array} \right\} \text{unabhängige Zufallsgrößen} \begin{cases} \text{je mit Verteilungsfunktion } F \\ \text{je mit Verteilungsfunktion } G. \end{cases}$$

Wir betrachten die Testprobleme

$$H_0: F \equiv G \quad (\text{d.h. beide Stichproben haben die gleiche Verteilung});$$

$$H_1: \text{a) } F \stackrel{\text{st}}{>} G \quad (\text{d.h. die } X_i \text{ sind stochastisch größer als die } Y_j);$$

$$\text{b) } F \stackrel{\text{st}}{<} G \quad (\text{d.h. die } X_i \text{ sind stochastisch kleiner als die } Y_j);$$

$$\text{c) } F \stackrel{\text{st}}{\neq} G \quad (\text{d.h. die } X_i \text{ sind stochastisch verschieden von den } Y_j).$$

Testgröße

Unter der Alternative  $H_1: F \stackrel{\text{st}}{>} G$  gilt  $P\{X > Y\} > P\{X < Y\}$  während unter der Nullhypothese aus Symmetriegründen  $P\{X > Y\} = P\{X < Y\}$ . Wir betrachten die Statistiken

$$U_+ = \text{Anzahl der Paare } (i, j) \text{ mit } X_i > Y_j$$

$$U_- = \text{Anzahl der Paare } (i, j) \text{ mit } X_i < Y_j$$

$$U_0 = \text{Anzahl der Paare } (i, j) \text{ mit } X_i = Y_j$$

Der Test von Wilcoxon-Mann-Whitney verwendet die Testgröße  $U = U_+ + \frac{1}{2}U_0$ . Diese Testgröße hängt zusammen mit der Rangsumme  $T_x$  (vgl. unten):  $T_x = U + \frac{1}{2}n_1(n_1 + 1)$ .

## 3. Test von Wilcoxon im stetigen Fall

Wir nehmen hier an, dass die beiden Verteilungsfunktionen  $F$  und  $G$  stetig sind; es gilt dann

$P\{X_i = Y_j\} = 0$  für alle  $i, j$ , und somit können wir davon ausgehen, dass keine Bindungen auftreten (engl.: tie = Bindung, Gleichstand).

Vorgehen:

- Gesamtstichprobe ordnen; Beobachtungen durch Rangzahlen ersetzen;
- unter  $H_0$  hat jede der  $\binom{n_1+n_2}{n_1}$  Permutationen die gleiche Wahrscheinlichkeit;
- als Testgröße verwendet der Test von Wilcoxon die Rangsumme  $T_x$  der  $X$ -Werte; die exakte Verteilung von  $T_x$  unter  $H_0$  ist berechenbar, sie hängt nicht von der speziellen Form von  $F (=G)$  ab; daher der Name "verteilungsfrei";
- Testvorschrift im Fall a):  
 Berechne die Rangsumme der  $x$ -Werte für die vorliegenden Stichproben  $\rightarrow t_{\text{obs}}$ ;  
 berechne den rechten  $p$ -Wert  $p_r = P_0\{T_x \geq t_{\text{obs}}\}$ ;  
 verwirf  $H_0$ , falls  $p_r \leq \alpha$ .

Analog im Fall b) und c).

Praktische Durchführung des Tests mit StatXact:

*Statistics, Two independent samples, Wilcoxon-Mann-Whitney, ...*

Beispiel:

$$n_1 = 2, \quad x_i = 5.0; 2.9$$

$$n_2 = 3, \quad y_j = 2.1; 4.5; 2.5$$

geordnete Gesamtstichprobe:  $2.1 < 2.5 < 2.9 < 4.5 < 5.0$

zugehörige Beob.namen:  $y_1 < y_3 < x_2 < y_2 < x_1$

zugehörige Stichprobe:  $y < y < x < y < x$

zugehörige Rangzahlen:  $1 \quad 2 \quad 3 \quad 4 \quad 5$

Die Rangsumme der x-Werte (1. Stichprobe) beträgt:  $T_x = 3 + 5 = 8 (= t_{\text{obs}})$ .

Die nachfolgende Tabelle listet die  $\binom{n_1+n_2}{n_1} = \binom{5}{2} = 10$  Permutationen von x- und y-Beobachtungen auf, welche unter  $H_0$  aus Symmetriegründen je die gleiche Wahrscheinlichkeit  $\frac{1}{10}$  besitzen.

Permutationen	Rangzahlen					$T_x$	$U$
	1	2	3	4	5		
1	x	x	y	y	y	3	0
2	x	y	x	y	y	4	1
3	x	y	y	x	y	5	2
4	x	y	y	y	x	6	3
5	y	x	x	y	y	5	2
6	y	x	y	x	y	6	3
7	y	x	y	y	x	7	4
8	y	y	x	x	y	7	4
9	y	y	x	y	x	8	5
10	y	y	y	x	x	9	6

Es gilt  $T_x = U + \frac{1}{2}n_1(n_1 + 1) = U + 3$ . Die Verteilung der Zufallsgröße  $T_x$  unter  $H_0$  ist somit gegeben durch die folgende Tabelle

$k$	3	4	5	6	7	8	9	
$P\{T_x = k\}$	0.1	0.1	0.2	0.2	0.2	0.1	0.1	1.0

und daraus ergibt sich, dass  $P\{T_x \geq t_{\text{obs}}\} = P\{T_x \geq 8\} = 0.2 (= p_r)$ . Da dieser p-Wert nicht klein genug ist, kann die Nullhypothese nicht abgelehnt werden.

#### 4. Test von Wilcoxon im allgemeinen Fall

Falls die Verteilungsfunktionen  $F$  und  $G$  Sprungstellen besitzen können, so können Bindungen auftreten, z.B.  $x_2 = y_7$  oder  $x_1 = y_3 = y_5$ . Die Beobachtungen einer Bindung müssen aus Symmetriegründen die gleiche Rangzahl erhalten, und damit hängt der Wertebereich und die Verteilung der Rangsumme von der Anzahl und Länge der Bindungen ab. Die Lage und die Länge der Bindungen hängt dabei ab von der Lage der Sprungstellen und den Sprunghöhen bei den beiden Verteilungsfunktionen. Es scheint, dass eine exakte Lösung ohne Zusatzannahmen über die Verteilungsfunktion nicht möglich ist. Dies ist jedoch nicht richtig, der Test von Wilcoxon kann ohne irgendwelche Annahmen über die Verteilungsfunktionen korrekt durchgeführt werden (datengesteuerter Test bei gegebenem Bindungsmuster).

### Bindungsmuster

Es sei  $n_1 = 2$  und  $n_2 = 3$ , d.h.  $n = n_1 + n_2 = 5$ . Wir bezeichnen die Beobachtungswerte der geordneten Gesamtstichprobe mit  $z_1 \leq z_2 \leq \dots \leq z_5$ . Nun könnte beispielsweise das folgende Beobachtungsmuster (Bindungsmuster) vorliegen:  $z_1 = z_2 < z_3 < z_4 < z_5$ ; dieses Muster ist charakterisiert mit  $(2,1,1,1)$ , d.h. die kleinsten beiden Werte bilden eine 2-er-Bindung, die übrigen drei Werte sind ungebunden. Ein anderes Bindungsmuster wäre  $z_1 < z_2 = z_3 < z_4 = z_5$ ; dieses Muster ist charakterisiert durch  $(1,2,2)$ . Jede Zerlegung  $(b_1, b_2, \dots)$  der natürlichen Zahl  $n = 5$  in Summanden mit  $b_1 + b_2 + \dots = n$  und  $b_i \geq 1$  liefert ein mögliches Bindungsmuster. Die Gesamtzahl der möglichen Bindungsmuster  $(b_1, b_2, \dots)$  bei gegebenem  $n$  beträgt  $2^{n-1}$  ( $n-1$  Positionen für ein mögliches Gleichheitszeichen). Die Wahrscheinlichkeit, dass ein bestimmtes Bindungsmuster vorliegt hängt ab von der Lage und Höhe der Sprungstellen bei den unbekanntenen Verteilungsfunktionen  $F$  und  $G$ .

### Verteilung der Testgröße bei gegebenem Bindungsmuster

Wir nehmen an, dass die Nullhypothese vorliegt, dass also die Zufallsgrößen  $X_i$  und  $Y_j$  alle unabhängig sind je mit der gleichen unbekanntenen Verteilungsfunktion  $F$ ; die Verteilungsfunktion  $F$  kann Sprungstellen besitzen, und somit können Bindungen auftreten. Es sei wiederum  $n_1 = 2$  und  $n_2 = 3$ , und wir nehmen an dass das Bindungsmuster  $(1,2,1,1)$  vorliegt. Auf die beiden Beobachtungen innerhalb der 2-er-Bindung fallen die Rangzahlen 2 und 3, und somit erhalten beide Beobachtungen das arithmetische Mittel  $(2+3)/2 = 2.5$  als Rangzahl. Die möglichen Permutationen mit den zugehörigen Werten der Rangsumme  $T_x$  sind in der nachfolgenden Tabelle aufgelistet. Dabei unterscheiden sich z.B. die ersten beiden Permutationen nur durch die Reihenfolge innerhalb der 2-er-Bindung, und sie sind als Beobachtungsmuster nicht unterscheidbar. Jede der 10 Permutationen hat aus Symmetriegründen die Wahrscheinlichkeit  $\frac{1}{10}$ .

Permutationen	Rangzahlen					$T_x$	$U$
	1	2.5	2.5	4	5		
1	$x$	$x$	$y$	$y$	$y$	3.5	0.5
2	$x$	$y$	$x$	$y$	$y$	3.5	0.5
3	$x$	$y$	$y$	$x$	$y$	5	2
4	$x$	$y$	$y$	$y$	$x$	6	3
5	$y$	$x$	$x$	$y$	$y$	5	2
6	$y$	$x$	$y$	$x$	$y$	6.5	3.5
7	$y$	$x$	$y$	$y$	$x$	7.5	4.5
8	$y$	$y$	$x$	$x$	$y$	6.5	3.5
9	$y$	$y$	$x$	$y$	$x$	7.5	4.5
10	$y$	$y$	$y$	$x$	$x$	9	6

Auch hier gilt  $T_x = U + \frac{1}{2}n_1(n_1 + 1) = U + 3$ . Die Verteilung der Zufallsgröße  $T_x$  unter  $H_0$  ist somit gegeben durch die folgende Tabelle

$k$	3.5	5	6	6.5	7.5	9	
$P\{T_x = k\}$	0.2	0.2	0.1	0.2	0.2	0.1	1.0

Diese Verteilung unterscheidet sich von der Verteilung, die wir oben im Fall ohne Bindungen gefunden haben. Die exakte Verteilung der Rangsumme  $T_x$  unter  $H_0$  bei gegebenem Bindungsmuster  $(b_1, b_2, \dots)$  ist also nur abhängig vom Bindungsmuster, und sie kann für jedes Bindungsmuster berechnet werden.

Es sei  $B_n$  die Gesamtheit aller möglichen Bindungsmuster zu  $n$ :

$$B_n = \{ \mathbf{b} = (b_1, b_2, \dots) \mid (b_1, b_2, \dots) \text{ ist Bindungsmuster zu } n \}.$$

Für jedes Bindungsmuster  $\mathbf{b} = (b_1, b_2, \dots)$  kann die exakte bedingte Verteilung der Rangsumme  $T_x$  gegeben  $\mathbf{b}$  bestimmt werden; mit Hilfe dieser Verteilung kann ein exakter Test zum Signifikanzniveau  $\alpha$  durchgeführt werden, und es gilt dann für diesen bedingten Test unter  $H_0: F \equiv G$

$$P_F \{ \text{Test verwirft } H_0 \mid \mathbf{b} = (b_1, b_2, \dots) \} \leq \alpha.$$

Somit gilt für den globalen Test (unbedingten Test):

$$\begin{aligned} P_F \{ \text{Test verwirft } H_0 \} &= \sum_{\mathbf{b} \in B_n} P_F(\mathbf{b}) \underbrace{P_F \{ \text{Test verwirft } H_0 \mid \mathbf{b} = (b_1, b_2, \dots) \}}_{\leq \alpha \text{ für jedes } \mathbf{b} \in B_n} \\ &\leq \alpha \underbrace{\sum_{\mathbf{b} \in B_n} P_F(\mathbf{b})}_{=1} = \alpha. \end{aligned}$$

Auf diese Weise kann der Test von Wilcoxon exakt durchgeführt werden auch wenn beliebige Bindungen vorliegen. Mit Hilfe von randomisierten Entscheidungen könnte das vorgegebene Signifikanzniveau  $\alpha$  bei jedem bedingten Test und damit auch beim globalen Test voll ausgeschöpft werden (wie z.B. beim exakten Vier-Felder-Test von Fisher).

Praktische Durchführung des Tests mit StatXact:

*Statistics, Two independent samples, Wilcoxon-Mann-Whitney, ...*

Die Prozedur berechnet zum Bindungsmuster, das durch die vorliegenden Daten gegeben ist, den exakten p-Wert beim gegebenen Bindungsmuster (bedingter oder datengesteuerter Test).

*Bemerkungen:*

a) Das Testproblem zum Zweistichproben-Test von Wilcoxon haben wir sehr allgemein formuliert als  $H_0: F \equiv G$  gegen  $H_1: F \not\equiv G$ , wobei über die Verteilungsfunktionen  $F$  und  $G$  keinerlei weitere Annahmen benötigt werden. Der Test kann exakt durchgeführt werden (z.B. mit StatXact), und der Test ist konsistent, d.h. wenn für ein festes Paar  $(F, G)$  die Alternative wahr ist, dann konvergiert die Ablehnwahrscheinlichkeit für  $n_1, n_2 \rightarrow \infty$  gegen 1. Dies ist sehr plausibel, denn wenn wir mit

$$U_+ = \text{Anzahl der Paare } (i, j) \text{ mit } X_i > Y_j$$

$$U_- = \text{Anzahl der Paare } (i, j) \text{ mit } X_i < Y_j$$

$$U_0 = \text{Anzahl der Paare } (i, j) \text{ mit } X_i = Y_j$$

bezeichnen, und mit  $U = U_+ + \frac{1}{2}U_0$ , dann gilt unter  $H_0$   $E_0(U) = n/2$ , während unter  $H_1$

$E_1(U) = np_1$  mit  $p_1 > 1/2$ . Man muß allerdings berücksichtigen, dass nicht alle Paare  $(X_i, Y_j)$  stochastisch unabhängig sind.

b) Ein Spezialfall unseres Testproblems sind die Lagemodelle mit  $G(x) = F(x - \mu)$ . Das Testproblem lautet dann  $H_0: \mu = 0$  gegen  $H_1: \mu < 0$ , denn für  $\mu < 0$  ist offensichtlich  $G(x) \geq F(x) \forall x \in \mathbb{R}$  und somit  $F \not\equiv G$ . Falls  $F$  eine Normalverteilung  $N(\mu, \sigma^2)$  ist, so gelangen wir zum Testproblem des klassischen doppelten t-Tests. Man kann zeigen, dass dann die asymptotische relative Effizienz des Tests von Wilcoxon im Vergleich zum doppelten t-Test  $3/\pi = 0.955$  beträgt. Dies ist ein erstaunlich hoher Wert, wenn man bedenkt, dass der Test von Wilcoxon einen viel breiteren Anwendungsbereich besitzt als der doppelte t-Test!

c) Wir betrachten das folgende Testproblem

$$\left. \begin{array}{l} X_1, \dots, X_{n_1} \\ Y_1, \dots, Y_{n_2} \end{array} \right\} \text{unabhängige Zufallsgrößen} \begin{cases} \text{je mit Verteilungsfunktion } F \\ \text{je mit Verteilungsfunktion } G. \end{cases}$$

$$\mu_F = \text{Median von } F$$

$$\mu_G = \text{Median von } G$$

$$H_0: \mu_F = \mu_G \quad \text{gegen} \quad H_1: \mu_F > \mu_G.$$

Man könnte meinen, dass auch dieses Problem mit dem Test von Wilcoxon gelöst werden kann. Dies ist jedoch nicht richtig, da unter der Alternative sowohl  $F \stackrel{\text{st}}{>} G$  als auch  $F \stackrel{\text{st}}{<} G$  sein kann. Das folgende Beispiel zeigt, dass  $\mu_F > \mu_G \not\Rightarrow F \stackrel{\text{st}}{>} G$ .

*Beispiel:*

$X$  und  $Y$  seien unabhängige Zufallsgrößen, wobei  $X$  eine Gleichverteilung in  $[0,1]$  besitzt, und die Dichte von  $Y$  gegeben ist durch

$$g(x) = \begin{cases} 1.5 & \text{für } 0 \leq x \leq 0.4 \\ 0 & \text{für } 0.4 < x < 1 \\ 1 & \text{für } 1 \leq x \leq 1.4 \\ 0 & \text{sonst.} \end{cases}$$

$F$  und  $G$  seien die Verteilungsfunktionen von  $X$  und  $Y$ . Dann gilt

$$\mu_F = \text{Median von } F = \frac{1}{2}$$

$$\mu_G = \text{Median von } G = \frac{1}{3}$$

und

$$P\{X < Y\} = \int_0^{1.4} F(y)g(y)dy = 0.52.$$

In diesem Beispiel ist also  $\mu_F > \mu_G$  und  $P\{X < Y\} > P\{X > Y\}$ , d.h.  $F \stackrel{\text{st}}{>} G$ .

Anstelle des Tests von Wilcoxon ist beim obigen Testproblem der Median-Test zum Zwei-Stichprobenproblem anzuwenden.