Data Mining / Intelligente Datenanalyse WS 2017/2018
PD Dr.-Ing. habil. Christian Borgelt

Lösung des 2. Übungsblattes

Aufgabe 7 Hauptkomponentenanalyse

Um die Hauptkomponentenanalyse durchzuführen, berechnen wir zunächst

$$\sum_{i=1}^{10} x_i = 30, \qquad \sum_{i=1}^{10} y_i = 30, \qquad \sum_{i=1}^{10} x_i^2 = 126, \qquad \sum_{i=1}^{10} y_i^2 = 120, \qquad \sum_{i=1}^{10} x_i y_i = 120.$$

Also sind die Elemente der Kovarianzmatrix

$$\begin{aligned} s_x^2 &= \tfrac{1}{9}\left(126 - \tfrac{1}{10} \cdot 30 \cdot 30\right) &&= 4, \\ s_y^2 &= \tfrac{1}{9}\left(120 - \tfrac{1}{10} \cdot 30 \cdot 30\right) &&= \tfrac{10}{3}, \\ s_{xy} &= \tfrac{1}{9}\left(120 - \tfrac{1}{10} \cdot 30 \cdot 30\right) &&= \tfrac{10}{3}. \end{aligned}$$

Der Korrelationskoeffizient ist folglich $r = \frac{s_{xy}}{s_x s_y} = \sqrt{\frac{5}{6}}$ und die Kovarianzmatrix

$$\mathbf{\Sigma} = \begin{pmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{pmatrix} = \begin{pmatrix} 4 & \frac{10}{3} \\ \frac{10}{3} & \frac{10}{3} \end{pmatrix}.$$

Von dieser Kovarianzmatrix gilt es nun, Eigenwerte und Eigenvektoren zu bestimmen. Gesucht sind also Vektoren $\vec{v}_1$ und $\vec{v}_2$ sowie zugehörige Werte λ_1 und λ_2, für die gilt

$$\mathbf{\Sigma}\vec{v}_k = \lambda_k \vec{v}_k, \quad k = 1, 2.$$

Wir wählen den Weg über das charakteristische Polynom, also

$$|\mathbf{\Sigma} - \lambda \mathbf{E}| = (4 - \lambda)\left(\frac{10}{3} - \lambda\right) - \frac{10}{3}\frac{10}{3} = \lambda^2 - \frac{22}{3}\lambda + \frac{20}{9}.$$

Die Nullstellen dieses Polynoms sind die gesuchten Eigenwerte. Es ist

$$\lambda_{1,2} = \frac{11}{3} \pm \sqrt{\left(\frac{11}{3}\right)^2 - \frac{20}{9}} = \frac{11 \pm \sqrt{101}}{3}.$$

Die Eigenwerte sind folglich $\lambda_1 \approx 7.02$ und $\lambda_2 \approx 0.32$. Die Eigenvektoren bestimmt man durch Einsetzen dieser Werte in die Gleichung $\mathbf{\Sigma}\vec{v}_k = \lambda_k \vec{v}_k$, $k = 1, 2$, und Lösen des sich ergebenden (unterbestimmten) Gleichungssystems. Man erhält

$$\vec{v}_1 = \left(\frac{10}{1 - \sqrt{101}}, -1\right)^T \approx (-1.105, -1)^T, \qquad \vec{v}_2 = \left(\frac{10}{1 + \sqrt{101}}, -1\right)^T \approx (0.905, -1)^T.$$

Aufgabe 8 Schätzfunktionen

Zur Bestimmung der Wahrscheinlichkeitsdichte des Schätzers W_2 gehen wir im Prinzip genauso vor, wie wir in der Vorlesung zur Bestimmung der Wahrscheinlichkeitsdichte des Schätzers für den Parameter θ einer Gleichverteilung auf dem Intervall $[0, \theta]$ vorgegangen sind. Es ist

$$\begin{aligned}
f_{W_2}(w;\theta) &= \frac{\mathrm{d}}{\mathrm{d}w} F_{W_2}(w;\theta) \;=\; \frac{\mathrm{d}}{\mathrm{d}w} P(W_2 \le w) \\
&= \frac{\mathrm{d}}{\mathrm{d}w} P(n \min\{X_1, \ldots, X_n\} \le w) \\
&= \frac{\mathrm{d}}{\mathrm{d}w} \left(1 - P\left(\min\{X_1, \ldots, X_n\} > \frac{w}{n}\right)\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}w} \left(1 - P\left(\bigwedge\nolimits_{i=1}^{n} X_i > \frac{w}{n}\right)\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}w} \left(1 - \prod_{i=1}^{n} P\left(X_i > \frac{w}{n}\right)\right) \\
&= -\frac{\mathrm{d}}{\mathrm{d}w} \left(1 - F_X\left(\frac{w}{n};\theta\right)\right)^n \\
&= \left(1 - F_X\left(\frac{w}{n};\theta\right)\right)^{n-1} \cdot f_X\left(\frac{w}{n},\theta\right)
\end{aligned}$$

mit

$$F_X(x;\theta) = \int_0^x f_X(t;\theta)\,\mathrm{d}t = 1 - e^{-\frac{x}{\theta}}.$$

Damit folgt

$$f_{W_2}(w;\theta) \;=\; \left(1 - \left(1 - e^{-\frac{w}{n\theta}}\right)\right)^{n-1} \frac{1}{\theta} e^{-\frac{w}{n\theta}} \;=\; \frac{1}{\theta} e^{-\frac{(n-1)w}{n\theta}} e^{-\frac{w}{n\theta}} \;=\; \frac{1}{\theta} e^{-\frac{w}{\theta}}.$$

Der Schätzer W_2 hat also genau die gleiche Wahrscheinlichkeitsdichte wie die einzelnen Zufallsvariablen. Da eine exponentialverteilte Zufallsvariable den Erwartungswert θ hat, wissen wir damit unmittelbar, daß dieser Schätzer erwartungstreu ist.

Aufgabe 9 Eigenschaften von Schätzfunktionen

Wir betrachten (siehe Hinweis in der Aufgabenstellung) die Zufallsvariable $W = \frac{1}{n} \sum_{i=1}^{n} Y_i$, die als arithmetisches Mittel der n Zufallsvariablen $Y_1, \ldots, Y_n$ mit

$$Y_i = \begin{cases} 1, & \text{falls das Ereignis } A \text{ im } i\text{-ten Versuch eintritt,} \\ 0, & \text{sonst,} \end{cases}$$

berechnet wird. Die Zufallsvariable W beschreibt offenbar die relative Häufigkeit r_A des Eintretens eines Ereignisses A in einer Stichprobe vom Umfang n für eine Binomialverteilung mit dem Parameter $p = P(A)$. Die Zufallsvariablen Y_i haben den Erwartungswert p, da

$$E(Y_i) = 0 \cdot (1 - P(A)) + 1 \cdot P(A) = 0 \cdot (1 - p) + 1 \cdot p = p,$$

und die Varianz $p(1 - p)$, da

$$\begin{aligned}
D^2(Y_i) &= E((Y_i - E(Y_i))^2) \;=\; E((Y_i - p)^2) \\
&= (1-p) \cdot (0-p)^2 + p \cdot (1-p)^2 \\
&= p^2 - p^3 + p - 2p^2 + p^3 \\
&= p - p^2 \;=\; p(1-p).
\end{aligned}$$

Durch Ausnutzen der Linearität des Erwartungswertes und der Addivität der Erwartungswerte bei einer Summe von Zufallsvariablen erhält man

$$E(W) \;=\; E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \;=\; \frac{1}{n}\sum_{i=1}^{n} E(Y_i) \;=\; \frac{1}{n}\sum_{i=1}^{n} p \;=\; \frac{1}{n}np \;=\; p.$$

Analog kann man für die Varianz herleiten, daß

$$D^2(W) = D^2\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(Y_i) = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{1}{n^2}np(1-p) \;=\; \frac{p(1-p)}{n}.$$

Der zweite Schritt der Umformung ist möglich, da die Zufallsvariablen Y_i paarweise unabhängig sind. (Beachte: Während die Addivität der Erwartungswerte stets gilt, wenn nur die Erwartungswerte existieren, muß für die Additivität der Varianzen außerdem paarweise stochastische Unabhängigkeit vorliegen.)

Da $E(W) = p$ gilt, ist W ein erwartungstreuer Schätzer für p, und wegen

$$\lim_{n\to\infty} D^2(W) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0$$

ist W auch ein konsistenter Schätzer (da die Varianz für $n \to \infty$ gegen 0 geht, konvergiert der Schätzer gegen den Wert des Parameters p).

Aufgabe 10 Erwartungstreue von Schätzfunktionen

a) Wir wissen aus der Aufgabenstellung, daß die Schätzer W_1 und W_2 erwartungstreu sind, daß also gilt: $E(W_1) = \theta$ und $E(W_2) = \theta$. Außerdem wissen wir, daß der Erwartungswert unempfindlich ist gegenüber linearen Transformationen, daß also gilt $E(aX+b) = aE(X)+b$, und daß der Erwartungswert einer Summe von Zufallsvariablen gleich der Summe der Erwartungswerte dieser Zufallsvariablen ist, daß also gilt $E(X+Y) = E(X)+E(Y)$. Folglich gilt

$$E(W) = E(aW_1 + bW_2) = E(aW_1) + E(bW_2) = aE(W_1) + bE(W_2) = a\theta + b\theta = (a+b)\theta.$$

Damit der Schätzer W erwartungstreu ist, muß $E(W) = \theta$ gelten. Damit folgt $(a+b)\theta = \theta$, also $a + b = 1$.

b) Die Aussage dieser Teilaufgabe ist eine direkte Verallgemeinerung der Teilaufgabe a) auf mehrere Schätzer. Es ist offenbar

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu.$$

In genau analoger Weise haben wir übrigens auch schon in Aufgabe 14 gerechnet.

Aufgabe 11 Effizienz von Schätzfunktionen

Um herauszufinden, welcher der beiden Schätzer effizienter ist, müssen wir die Varianzen der beiden Schätzer berechnen und vergleichen. Wir beginnen mit dem Schätzer W_1. Aus der Vorlesung wissen wir, daß die Wahrscheinlichkeitsdichte des Schätzers W_1

$$f_{W_1}(w;\theta) = \frac{n^{n+1}}{(n+1)^n}\frac{w^{n-1}}{\theta^n}, \qquad 0 \le w \le \frac{n+1}{n}\theta,$$

ist. Um die Varianz dieses Schätzers zu berechnen, nutzen wir die Beziehung

$$D^2(W_1) = E(W_1^2) - (E(W_1))^2 = E(W_1^2) - \theta^2.$$

(Daß $E(W_1) = \theta$ ist, wissen wir, weil wir den Schätzer W_1 in der Vorlesung als erwartungstreu für den Parameter θ nachgewiesen haben.) Wir müssen folglich nur noch $E(W_1^2)$ berechnen:

$$\begin{aligned}
E(W_1^2) &= \int_0^{\frac{n+1}{n}\theta} \frac{n^{n+1}}{(n+1)^n} \frac{w^{n-1}}{\theta^n} w^2 \,\mathrm{d}w \\
&= \frac{n^{n+1}}{(n+1)^n\theta^n} \int_0^{\frac{n+1}{n}\theta} w^{n+1} \,\mathrm{d}w \\
&= \frac{n^{n+1}}{(n+1)^n\theta^n} \left[\frac{w^{n+2}}{n+2}\right]_0^{\frac{n+1}{n}\theta} \\
&= \frac{n^{n+1}}{(n+1)^n\theta^n} \frac{(n+1)^{n+2}}{(n+2)n^{n+2}} \theta^{n+2} \quad = \quad \frac{(n+1)^2}{n(n+2)}\theta^2.
\end{aligned}$$

Also ist

$$D^2(W_1) \quad = \quad \frac{(n+1)^2}{n(n+2)}\theta^2 - \theta^2 \quad = \quad \frac{\theta^2}{n(n+2)}.$$

Betrachten wir nun den Schätzer W_2. Zunächst müssen wir seine Wahrscheinlichkeitsdichte bestimmen, wozu wir genauso vorgehen wie in der Vorlesung zur Bestimmung der Wahrscheinlichkeitsdichte des Schätzers W_1 und in Aufgabe 13. Es ist

$$\begin{aligned}
f_{W_2}(w;\theta) &= \frac{\mathrm{d}}{\mathrm{d}w} F_{W_2}(w;\theta) \ = \ \frac{\mathrm{d}}{\mathrm{d}w} P(W_2 \le w) \\
&= \frac{\mathrm{d}}{\mathrm{d}w} P((n+1)\min\{X_1,\ldots,X_n\} \le w) \\
&= \frac{\mathrm{d}}{\mathrm{d}w}\left(1 - P\left(\min\{X_1,\ldots,X_n\} > \frac{w}{n+1}\right)\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}w}\left(1 - P\left(\bigwedge\nolimits_{i=1}^n X_i > \frac{w}{n+1}\right)\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}w}\left(1 - \prod_{i=1}^n P\left(X_i > \frac{w}{n+1}\right)\right) \\
&= -\frac{\mathrm{d}}{\mathrm{d}w}\left(1 - F_X\left(\frac{w}{n+1};\theta\right)\right)^n \\
&= \frac{n}{n+1} \cdot \left(1 - F_X\left(\frac{w}{n+1};\theta\right)\right)^{n-1} f_X\left(\frac{w}{n+1},\theta\right)
\end{aligned}$$

mit

$$F_X(x;\theta) \ = \ \int_{-\infty}^x f_X(t;\theta)\,\mathrm{d}t \ = \ \begin{cases} 0, & \text{falls } x \le 0, \\ \frac{x}{\theta}, & \text{falls } 0 \le x \le \theta, \\ 1, & \text{falls } x \ge \theta. \end{cases}$$

Damit folgt

$$f_{W_2}(w;\theta) = \frac{n}{(n+1)\theta}\left(1 - \frac{w}{(n+1)\theta}\right)^{n-1}, \qquad 0 \le w \le (n+1)\theta.$$

Auch dieser Schätzer ist erwartungstreu für den Parameter θ, denn

$$\begin{aligned}
E(W_2) &= \int_0^{(n+1)\theta} \frac{n}{(n+1)\theta}\left(1-\frac{w}{(n+1)\theta}\right)^{n-1} w\,\mathrm{d}w \\
&= \left[-\left(1-\frac{w}{(n+1)\theta}\right)^{n} w\right]_0^{(n+1)\theta} + \int_0^{(n+1)\theta}\left(1-\frac{w}{(n+1)\theta}\right)^{n}\mathrm{d}w \\
&= 0+\left[-\theta\left(1-\frac{w}{(n+1)\theta}\right)^{n+1}\right]_0^{(n+1)\theta} \quad = \quad \theta.
\end{aligned}$$

Da wir natürlich wieder

$$D^2(W_2) = E(W_2^2) - (E(W_2))^2 = E(W_2^2) - \theta^2$$

nutzen, bleibt uns wieder nur $E(W_2^2)$ zu berechnen:

$$\begin{aligned}
E(W_2^2) &= \int_0^{(n+1)\theta} \frac{n}{(n+1)\theta}\left(1-\frac{w}{(n+1)\theta}\right)^{n-1} w^2\,\mathrm{d}w \\
&= \left[-\left(1-\frac{w}{(n+1)\theta}\right)^{n} w^2\right]_0^{(n+1)\theta} + \int_0^{(n+1)\theta}\left(1-\frac{w}{(n+1)\theta}\right)^{n} 2w\,\mathrm{d}w \\
&= 0+2\int_0^{(n+1)\theta}\left(1-\frac{w}{(n+1)\theta}\right)^{n} w\,\mathrm{d}w \\
&= 2\left[-\theta\left(1-\frac{w}{(n+1)\theta}\right)^{n+1} w\right]_0^{(n+1)\theta} + 2\int_0^{(n+1)\theta}\theta\left(1-\frac{w}{(n+1)\theta}\right)^{n+1}\mathrm{d}w \\
&= 0+2\left[-\frac{(n+1)\theta^2}{n+2}\left(1-\frac{w}{(n+1)\theta}\right)^{n+2}\right]_0^{(n+1)\theta} \\
&= \frac{2(n+1)}{n+2}\theta^2
\end{aligned}$$

(Man beachte die zweimalige Anwendung der partiellen Integration in dieser Rechnung.)
Also ist

$$D^2(W_2) \quad = \quad \frac{2(n+1)}{n+2}\theta^2 - \theta^2 \quad = \quad \frac{n\theta^2}{n+2}.$$

Da nun

$$D^2(W_1) = \frac{\theta^2}{n(n+2)} \quad < \quad \frac{n\theta^2}{n+2} = D^2(W_2)$$

für $n > 1$, ist der Schätzer W_1 effizienter. In der Tat ist die Überlegenheit des Schätzer W_1 sehr ausgeprägt: Seine relative Effizienz im Vergleich zum Schätzer W_2 ist

$$\frac{D^2(W_2)}{D^2(W_1)} \quad = \quad \frac{n\theta^2}{n+2}\,\frac{n(n+2)}{\theta^2} \quad = \quad n^2.$$

Aufgabe 12 Maximum-Likelihood-Schätzung

Die Wahrscheinlichkeit einer Stichprobe vom Umfang n einer Gleichverteilung auf dem Intervall $[0,\theta]$ ist

$$L(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f_X(x_n;\theta) = \left(\frac{1}{\theta}\right)^n.$$

Der formale Weg, diese Funktion nach θ abzuleiten und das Ergebnis 0 zu setzen, führt hier jedoch nicht zum Ziel, denn so erhält man

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1}{\theta}\right)^n = \frac{\mathrm{d}}{\mathrm{d}\theta}\theta^{-n} = -n\theta^{-n-1} \stackrel{!}{=} 0,$$

also $\hat{\theta} = 0$. Dies ist aber offenbar kein sinnvoller Schätzer. Wir müssen vielmehr berücksichtigen, daß als Nebenbedingung einzuhalten ist, daß alle aufgetretenen Stichprobenwerte auch möglich sind. Wir müssen also einen Wert θ bestimmen, für den $\theta \geq \max\{x_1, \ldots, x_n\}$ gilt. Außerdem sollte θ so klein wie möglich sein, damit $\frac{1}{\theta}$ so groß wie möglich und folglich die Stichprobe maximal wahrscheinlich ist. Damit erhalten wir als Schätzer

$$\hat{\theta} = \max\{x_1, \ldots, x_n\},$$

also den Schätzer, den wir zuerst in der Vorlesung betrachtet haben. Wie in der Vorlesung abgeleitet, ist dieser Schätzer konsistent, aber nicht erwartungstreu.

Aufgabe 13 Maximum-Likelihood-Schätzung

Die Wahrscheinlichkeit dafür, daß in einer Stichprobe vom Umfang n ein Ereignis A x-mal eintritt, wird durch eine Binomialverteilung mit dem Parameter $\theta = P(A)$ beschrieben:

$$L(x; \theta) = b_X(x; \theta, n) = \tbinom{n}{x}\theta^x(1-\theta)^{n-x}.$$

Ableiten dieser Likelihood-Funktion nach dem zu schätzenden Parameter θ liefert:

$$\begin{aligned}\frac{\partial L}{\partial \theta} &= \tbinom{n}{x}(x\theta^{x-1}(1-\theta)^{n-x} - \theta^x(n-x)(1-\theta)^{n-x-1}) \\ &= \tbinom{n}{x}\theta^{x-1}(1-\theta)^{n-x-1}(x(1-\theta) - \theta(n-x)) \quad \stackrel{!}{=} \quad 0\end{aligned}$$

Wir erhalten folglich

$$x(1-\hat{\theta}) = \hat{\theta}(n-x) \quad \Leftrightarrow \quad x - x\hat{\theta} = n\hat{\theta} - x\hat{\theta} \quad \Leftrightarrow \quad \hat{\theta} = \frac{x}{n}.$$

Für eine Polynomialverteilung ist die Ableitung analog. Wir haben die Likelihood-Funktion

$$L(x_1, \ldots, x_k; \theta_1, \ldots, \theta_{k-1}) = \frac{n!}{\prod_{i=1}^k x_i!}\left(\prod_{i=1}^{k-1}\theta_i^{x_i}\right)\left(1 - \sum_{i=1}^{k-1}\theta_i\right)^{x_k}$$

zu betrachten, wobei die $\theta_i = P(A_i)$ die Wahrscheinlichkeiten der k verschiedenen Ereignisse A_i sind. Die x_i bezeichnen die (absoluten) Häufigkeiten, mit denen die verschiedenen Ereignisse A_i in einer vorliegenden Stichprobe aufgetreten sind. Man beachte, daß nur die Wahrscheinlichkeiten θ_1 bis θ_{k-1} als Parameter auftreten, da ja $\sum_{i=1}^k \theta_i = 1$ gelten muß und folglich mit θ_1 bis θ_{k-1} bereits alle Wahrscheinlichkeiten festliegen.

Durch Ableiten der Likelihood-Funktion nach den Parametern erhält man

$$\begin{aligned}\frac{\partial L}{\partial \theta_j} &= \frac{n!}{\prod_{i=1}^k x_i!} x_j\theta_j^{x_j-1}\left(\prod_{i=1, i\neq j}^{k-1}\theta_i^{x_i}\right)\left(1 - \sum_{i=1}^{k-1}\theta_i\right)^{x_k} - \left(\prod_{i=1}^{k-1}\theta_i^{x_i}\right)x_k\left(1 - \sum_{i=1}^{k-1}\theta_i\right)^{x_k-1} \\ &= \frac{n!}{\prod_{i=1}^k x_i!}\left(\prod_{i=1, i\neq j}^{k-1}\theta_i^{x_i}\right)\theta_j^{x_j-1}\left(1 - \sum_{i=1}^{k-1}\theta_i\right)^{x_k-1}\left(x_j\left(1 - \sum_{i=1}^{k-1}\theta_i\right) - \theta_j x_k\right) \quad \stackrel{!}{=} \quad 0.\end{aligned}$$

Daher muß gelten (der letzte Faktor der obigen Gleichung muß stets 0 sein)

$$\forall j, 1 \leq j < k: \qquad x_j \left(1 - \sum_{i=1}^{k-1} \hat{\theta}_i\right) = \hat{\theta}_j x_k.$$

Dieses Gleichungssystem hat die Lösung $\forall i, 1 \leq j < k : \hat{\theta}_j = \frac{x_j}{n}$.

Formal etwas einfacher kann man u.U. das Gleichungssystem mit Hilfe der Methode der sogenannten *Lagrange-Multiplikatoren* erhalten. In diesem Fall erweitert man die Likelihood-Funktion zu der sogenannten *Lagrange-Funktion*

$$L'(x_1, \ldots, x_k; \theta_1, \ldots, \theta_k, \lambda) = \frac{n!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} \theta_i^{x_i} + \lambda \left(1 - \sum_{i=1}^{k} \theta_i\right)$$

mit dem (unbekannten) Lagrange-Multiplikator λ, der wie die anderen Unbekannten $\theta_1, \ldots, \theta_k$ behandelt wird. (Genaueres zu dieser Methode in der Vorlesung im Abschnitt über Fuzzy-Clustering.) Für die Lagrange-Funktion gilt ebenfalls, daß alle partiellen Ableitungen verschwinden müssen, also

$$\frac{\partial L'}{\partial \theta_j} = \frac{n!}{\prod_{i=1}^{k} x_i!} x_j \theta_j^{x_j - 1} \left(\prod_{i=1, i \neq j}^{k-1} \theta_i^{x_i}\right) + \lambda \quad \stackrel{!}{=} \quad 0$$

und (Reproduktion der Nebenbedingung für die θ_j)

$$\frac{\partial L'}{\partial \lambda} = 1 - \sum_{i=1}^{k-1} \theta_i \quad \stackrel{!}{=} \quad 0.$$