Zeitschrift: Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker = Bulletin / Association des Actuaires Suisses = Bulletin / Association of Swiss Actuaries

Band: 58 (1958)

Artikel: Ein theoretischer Beitrag zur statistischen Erfassung der Gesamtbetriebsunfallkosten

Autor: Bühlmann, H.

DOI: https://doi.org/10.5169/seals-966797

Ein theoretischer Beitrag zur statistischen Erfassung der Gesamtbetriebsunfallkosten

Von *H. Bühlmann*, Berkeley [1])

1. Kapitel

Die statistische Erfassung der Gesamtbetriebsunfallkosten ist in letzter Zeit Gegenstand einiger Artikel in den «Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker» gewesen. Die vorliegende Arbeit nimmt das Thema diesmal in einer mehr theoretischen Form wieder auf. Der Grund für diese «graue Theorie» ist ein doppelter. Einerseits war der Schreibende durch praktische Versuche W. Hartmanns darauf aufmerksam geworden, wie viele sehr interessante theoretische Probleme die Betriebsunfallstatistik bietet. Andererseits bot sich ihm im Rahmen eines Projekts des Department of Statistics der University of California, Berkeley, die Möglichkeit, sich mit dem Fragenkomplex theoretisch auseinanderzusetzen.

Die Gesamtbetriebsunfallkosten sind eine Funktion der Unfallkosten und der Unfallhäufigkeit. Das 2. Kapitel gibt und diskutiert die mathematischen Hypothesen, welche wir über diese 2 stochastischen Grössen machen. Im 3. Kapitel wird die Verteilungsfunktion der Gesamtbetriebsunfallkosten abgeleitet. Das 4. Kapitel beschäftigt sich mit der Schätzung der Parameter. Auf Grund dieser Schätzungen ist es dann möglich, alle statistischen Masszahlen (z. B. Erwartungswert und Streuung) der Gesamtbetriebsunfallkosten zu berechnen. Das 5. Kapitel handelt von den Veränderungen in den Grundgesamtheiten der Unfallkosten resp. Unfallhäufigkeiten. Es wird in diesem Kapitel eine Methode zur Kontrolle der Parameter entwickelt, welche anzeigt, wenn eine neue Schätzung derselben vorgenommen werden muss.

[1]) Dieser Artikel wurde geschrieben mit Unterstützung des Office of Naval Research der Vereinigten Staaten von Amerika. Die Regierung der Vereinigten Staaten hat das Recht, den Artikel oder Teile desselben für ihre Zwecke zu reproduzieren.

2. Kapitel

Mathematische Hypothesen über die Grundgesamtheiten der Unfallkosten und Unfallhäufigkeiten

a) Grundgesamtheit der Unfallhäufigkeiten

Die Anzahl Unfälle, die in einem Betrieb während einer bestimmten Beobachtungsperiode gezählt werden, kann durch die Normalverteilung beschrieben werden, sofern der Betrieb genügend gross und die Beobachtungsperiode genügend lang ist. W. Hartmann hat durch praktische Versuche gezeigt, dass es statthaft ist, normalverteilte Unfallhäufigkeiten anzunehmen, sobald mehr als 14 Unfälle pro Beobachtungsperiode zu erwarten sind. Diese Voraussetzung sei im folgenden stillschweigend angenommen. Teilt man die Anzahl Unfälle durch die totale Anzahl der im Beobachtungsintervall geleisteten Arbeitsstunden, so haben wir ebenfalls eine Grösse vor uns, die normalverteilt ist. Wir definieren folgende stochastische Variable

$$w = \frac{k}{g},$$

wobei k = Anzahl Unfälle im Beobachtungsintervall,
g = total geleistete Arbeitsstunden im Beobachtungsintervall,

und machen die folgende statistische Hypothese: w ist normalverteilt mit Mittelwert ν und Varianz $\frac{\delta^2}{g}$. Diese Voraussetzung schreiben wir in symbolischer Form wie folgt:

Hypothese I: $$w \approx N\left(\nu, \frac{\delta^2}{g}\right).$$

Der Grund, dass die Varianz in der Form $\frac{\delta^2}{g}$ angenommen wurde, ist darin zu suchen, dass w ein Durchschnitt ist. Die Varianz eines Durchschnittes von g unabhängigen Elementen ist aber bekanntlich gleich der Varianz der Einzelbeobachtung dividiert durch die Anzahl der Elemente. Der Ansatz: Varianz $= \frac{\delta^2}{g}$ hat den Vorteil, dass die Grösse δ^2 unverändert bleibt, auch wenn sich g, die Anzahl der geleisteten Arbeitsstunden, verändert.

b) *Unfallkostengrundgesamtheit*

In einer früheren Arbeit wurde die logarithmische Normalverteilung als geeignete Verteilungsfunktion gefunden, um den Verlauf der Unfallkosten zu erklären. Der «Fit» der logarithmischen Normalverteilung ist um so besser, aus je mehr Unfällen die Unfallkosten entstanden sind. Auch hier hat W. Hartmann durch praktische Versuche festgestellt, dass das arithmetische Mittel der Unfallkosten aus mehr als 16 Fällen als log. normalverteilt angenommen werden darf. Aus rein theoretischen Gründen scheint es hingegen vorteilhafter, statt der log. Normalverteilung eine andere Verteilungsfunktion zu Hilfe zu nehmen, nämlich die Γ-Verteilung. Diese Verteilungsfunktion zeigt einen der log. Normalverteilung sehr verwandten Verlauf, eignet sich aber für die später vorzunehmenden Operationen viel besser. Die Γ-Verteilung ist durch folgende Dichtefunktion definiert:

$$f_{\Gamma(\alpha,a)}(x) = \frac{a^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-ax}, \quad \text{für } x > 0.$$

Die Eigenschaften dieser Verteilung sind die folgenden:

Mittelwert: $\frac{\alpha}{a}$, (1)

Streuung: $\frac{\alpha}{a^2}$, (2)

charakteristische Funktion: $\left(1 - \frac{iu}{a}\right)^{-\alpha}$. (3)

Aus der Form der charakteristischen Funktion liest man sofort ab, dass bei gleichem a die Faltung von 2 Γ-Verteilungen wieder eine Γ-Verteilung ergibt. Dies halten wir symbolisch in folgender Form fest:

$$F_{\Gamma(\alpha_1,a)}(x) * F_{\Gamma(\alpha_2,a)}(x) = F_{\Gamma(\alpha_1+\alpha_2,a)}(x). \tag{4}$$

Wenn x Γ-verteilt mit α, a, dann ist λx Γ-verteilt mit $\alpha, \frac{a}{\lambda}$. (5)

Als statistische Grösse für die Unfallkosten ist folgende stochastische Variable zweckmässig:

$$x = \frac{K}{kI},$$

K = Summe der Unfallkosten im Beobachtungsintervall,
k = Anzahl Unfälle im Beobachtungsintervall,
I = mittlerer Stundenlohn.

Bemerkung: Die Division durch I wurde vorgenommen, um die Variable von Indexschwankungen unabhängig zu machen.

Die mathematische Hypothese für die Variable x ist die folgende: x ist Γ-verteilt mit

$$\text{Mittelwert:} \quad \frac{\gamma}{c},$$

$$\text{Varianz:} \quad \frac{\gamma}{kc^2}.$$

Der Grund, weshalb in der Varianz k im Nenner auftritt, ist wiederum darin zu sehen, dass x aus einer Mittelbildung entstanden ist. Berechnet man die Parameter α und a der Γ-Verteilung aus dieser Annahme, so findet man

$$\alpha = k,$$
$$a = ck.$$

Symbolischerweise schreiben wir die Voraussetzung betreffend x deshalb in folgender Form

Hypothese II: $\quad x \approx \Gamma(k\gamma, kc)$

oder $\quad x \approx \Gamma(gw\gamma, gwc)$.

Die zwei Schreibweisen können leicht ineinander übergeführt werden, wenn man beachtet, dass

$$k = gw.$$

Die in diesem Kapitel beschriebenen Hypothesen bilden die Grundlage der im folgenden entwickelten Theorie. Die Voraussetzungen scheinen zweckmässig zu sein und mit empirischem Zahlenmaterial oft gut übereinzustimmen. Es bleibt aber im Einzelfalle der numerischen Analysis überlassen, über Annahme oder Verwerfung der 2 Hypothesen zu entscheiden.

3. Kapitel

Die Verteilungsfunktion der Gesamtbetriebsunfallkosten

Es ist in der Betriebsunfallversicherung üblich, folgende Masszahl für das Risiko zu benützen:

$$z = \frac{\text{Summe der Unfallkosten}}{\text{Summe der ausbezahlten Löhne}}.$$

Man überzeugt sich leicht, dass diese Variable z gerade das Produkt der im vorigen Kapitel beschriebenen Variabeln x und w darstellt.

$$x = \frac{K}{kI}, \qquad w = \frac{k}{g};$$

$$wx = \frac{kK}{kIg} = \frac{\text{Summe der Unfallkosten}}{\text{Lohnsumme}} = z.$$

Die Verteilungsfunktion von z ergibt sich folgendermassen:

$$H_z(\xi) = W(z < \xi) = W(xw < \xi) = \int W(xs < \xi \mid w = s)\, dW(w < s)$$
$$= \int W(gxs < g\xi \mid k = gs)\, dW(w < s) = \int F_{\Gamma(\gamma gs, c)}(g\xi)\, dW(w < s).$$

Die Begründung des letzten Schrittes ergibt sich direkt aus Eigenschaft (5) der Γ-Verteilung und Hypothese I. Unter Verwendung von Hypothese II finden wir

$$H_z(\xi) = \int F_{\Gamma(\gamma gs, c)}(g\xi)\, \varphi\left(\frac{s-\nu}{\delta}\sqrt{g}\right) d\left(\frac{s-\nu}{\delta}\sqrt{g}\right). \tag{6}$$

Die Dichtenfunktion ergibt sich durch Differenzieren:

$$h_z(\xi) = \int g f_{\Gamma(\gamma gs, c)}(g\xi)\, \varphi\left(\frac{s-\nu}{\delta}\sqrt{g}\right) d\left(\frac{s-\nu}{\delta}\sqrt{g}\right), \tag{7}$$

wobei φ die Dichte der normierten Normalverteilung darstellt.

Diese Verteilungsfunktion stellt einen recht verwickelten analytischen Ausdruck dar. Die numerische Auswertung desselben (z.B. zur Bestimmung von Quantilen) ist wohl sehr langwierig, andererseits aber unter Benutzung der Pearsonschen «Tables of the Incomplete Γ-function» möglich. Die Momente der Verteilung lassen sich hingegen sehr einfach bestimmen, und sofern der Versicherer nur an Erwartungswert und eventueller Streuung interessiert ist, kann die numerische Auswertung des Integrals beiseite gelassen werden. Man findet für die Momente:

$$E(z) = E(x)\,E(w) = \frac{\gamma}{c\nu}, \tag{8}$$

$$\operatorname{Var}(z) = \frac{\gamma^2\delta^2 + \gamma\delta}{g c^2}. \tag{9}$$

4. Kapitel

Schätzung der Parameter

Die Verteilungsfunktion $H_z(\xi)$ besitzt 4 unbekannte Parameter: γ, c, ν, δ^2. Zur Schätzung dieser Parameter nehmen wir an, dass folgende Information zur Verfügung steht.

w_i = Anzahl Unfälle pro Arbeitsstunde in der Beobachtungsperiode i,
g_i = Anzahl geleisteter Arbeitsstunden in der Beobachtungsperiode i,
x_i = Unfallkosten pro Unfall/mittlerer Stundenlohn in der Beobachtungsperiode i,
$k_i = g_i w_i$ = Anzahl Unfälle in der Beobachtungsperiode i.

Wir schätzen auf Grund des «Maximum-Likelihood»-Prinzips. Die «Maximum-Likelihood»-Funktion hat folgende Gestalt:

$$L = \frac{1}{2}\sum_{i=1}^{n}\log g_i - n\log\delta - \frac{1}{2}\sum_{i=1}^{n}\frac{g_i}{\delta^2}(w_i-\nu)^2 + \sum_{i=1}^{n}k_i\gamma\log k_i c$$
$$-\sum_{i=1}^{n}\log\Gamma(k_j\gamma) + \sum_{i=1}^{n}(k_i\gamma-1)\log x_i - \sum_{i=1}^{n}k_i c x_i. \tag{10}$$

Differenzieren wir nach den Parametern, so finden wir folgende zwei Gleichungssysteme:

a) für ν, δ^2: $\displaystyle\sum_{i=1}^{n} g_i(x_i-\nu) = 0$

$$-\frac{n}{\delta} + \frac{1}{\delta^3}\sum_{i=1}^{n} g_i(x_i-\nu)^2 = 0;$$

b) für γ, c: $\displaystyle\frac{1}{c}\sum_{i=1}^{n}k_i\gamma = \sum_{i=1}^{n}k_i x_i = 0$

$$\sum_{i=1}^{n}k_i(\log k_i c) - \sum_{i=1}^{n}\frac{\Gamma'(k_i\gamma)}{\Gamma(k_i\gamma)}k_i + \sum_{i=1}^{n}k_i\log x_i = 0.$$

Die Auflösung ergibt: (Maximum-Likelihood-Schätzungen seien von den «wahren» Parameterwerten durch einen Zirkumflex [^] unterschieden.)

$$\hat{\nu} = \frac{\sum_{i=1}^{n} g_i x_i}{\sum_{i=1}^{n} g_i}, \tag{11}$$

$$\hat{\delta}^2 = \frac{1}{n} \sum g_i (x_i - \hat{\nu})^2, \tag{12}$$

$$\frac{\hat{\gamma}}{\hat{c}} = \frac{\sum_{i=1}^{n} k_i x_i}{\sum_{i=1}^{n} k_i}. \tag{13}$$

$\hat{\gamma}$ kann bestimmt werden aus der Gleichung

$$\begin{gathered} \log \hat{\gamma} \sum_{i=1}^{n} k_i - \sum_{i=1}^{n} k_i \frac{\Gamma'(k_i \hat{\gamma})}{\Gamma(k_i \hat{\gamma})} = \\ = \log\left(\sum_{i=1}^{n} k_i x_i\right) \sum_{i=1}^{n} k_i - \log\left(\sum_{i=1}^{n} k_i\right) \sum_{i=1}^{n} k_i - \sum_{i=1}^{n} k_i \log k_i - \sum_{i=1}^{n} k_i \log x_i . \end{gathered} \tag{14}$$

Bemerkung: Diese letzte Beziehung kann numerisch mit Hilfe der Digammatafeln (K. Pearson, Cambridge University, 1919) gelöst werden.

Die Beziehungen (11)–(14) zeigen, dass die Maximum-Likelihood-Schätzungen von ν, δ^2 nur von den beobachteten Grössen aus der Grundgesamtheit der Unfallhäufigkeiten abhängen, während $\hat{\nu}$, $\hat{c}$ nur von den Beobachtungen aus der Kostengrundgesamtheit bestimmt werden. Diese Zweiteilung des Problems war vom «anschaulichen» Standpunkt aus vielleicht zu erwarten, mathematisch ist sie aber keineswegs trivial, da die Grundgesamtheiten der Unfallkosten und Unfallhäufigkeiten ja nicht als voneinander unabhängig vorausgesetzt wurden.

Der Versicherer, für den der Erwartungswert wohl die wichtigste statistische Grösse ist, wird nicht unbedingt daran interessiert sein, die besten Schätzungen für die Parameter ν, δ^2, γ, c zu kennen. Für ihn handelt es sich vielmehr darum, die beste Schätzung für den *Erwartungswert* $E(z)$ zu finden. Die Frage, welche er deshalb an dieser Stelle wohl aufwirft, wird folgende sein.

Wenn man in Formel (8) die Schätzwerte für die Parameter einsetzt, so erhält man

$$\widetilde{\widetilde{E(z)}} = \frac{\hat{\gamma}}{\hat{c}}\,\hat{\nu}\,.$$

Ist $\widetilde{\widetilde{E(z)}}$ die Maximum-Likelihood-Schätzung für $E(z)$? Um die Frage abzuklären, gehen wir zurück zu unserer «Maximum-Likelihood»-Funktion und führen den Erwartungswert

$$E(z) = \frac{\gamma}{c} = \mu$$

als neuen Parameter ein. Indem wir ν eliminieren, haben wir dann statt der 4 alten Parameter: ν, δ^2, γ, c 4 neue: μ, δ^2, γ, c, und die «Maximum-Likelihood»-Funktion hat folgende Gestalt:

$$L = \frac{1}{2}\sum_{i=1}^{n} \log g_i - n \log \delta - \frac{1}{2\delta^2}\sum_{i=1}^{n} g_i \left(w_i - \frac{\mu c}{\gamma}\right)^2 + \sum_{i=1}^{n} k_i \gamma \log (k_i c) -$$

$$- \sum_{i=1}^{n} \log \Gamma(k_i\gamma) + \sum_{i=1}^{n} (k_i\gamma - 1) \log x_i - \sum_{i=1}^{n} k_i c x_i\,. \quad (10^*)$$

Differenzieren nach μ ergibt:

$$\sum_{i=1}^{n} g_i \left(w_i - \frac{\mu c}{\gamma}\right) = 0 \tag{15}$$

und

$$\hat{\hat{\mu}} = \frac{\hat{\hat{\gamma}}}{\hat{\hat{c}}} \cdot \frac{\sum\limits_{i=1}^{n} g_i w_i}{\sum\limits_{i=1}^{n} g_i}\,. \tag{16}$$

Bemerkung: Die Schätzungen aus der Maximum-Likelihood-Funktion (10*) bezeichnen wir im Unterschiede zu den frühern Schätzwerten mit doppeltem Zirkumflex [$\hat{\hat{}}$].

Auf Grund von Beziehung (15) stellt man ohne Rechnung fest, dass das Gleichungssystem für $\hat{\hat{\gamma}}$ und $\hat{\hat{c}}$, welches man aus (10*) erhält, dasselbe sein wird wie das Gleichungssystem, welches aus (10) abgeleitet $\hat{\gamma}$ und $\hat{c}$ bestimmte. Das bedeutet, dass

$$\hat{\hat{\gamma}} = \hat{\gamma}\,, \tag{17}$$

$$\hat{\hat{c}} = \hat{c}\,, \tag{18}$$

somit
$$\hat{\hat{\mu}} = \frac{\hat{\gamma}}{\hat{c}}\,\hat{\nu}$$

oder
$$\hat{\hat{\mu}} = \frac{\sum\limits_{i=1}^{n} k_i x_i}{\sum\limits_{i=1}^{n} k_i} \cdot \frac{\sum\limits_{i=1}^{n} g_i w_i}{\sum\limits_{i=1}^{n} g_i} = \frac{\sum\limits_{i=1}^{n} k_i x_i}{\sum\limits_{i=1}^{n} g_i} \,. \tag{19}$$

Diese Formel stellt die Maximum-Likelihood-Schätzung für den Erwartungswert $E(z)$ dar.

5. Kapitel

Kontrolle der Parameter

Für unsere Schätzungen haben wir stillschweigend angenommen, dass die Grundgesamtheiten der Unfallkosten und -häufigkeiten sich im Verlaufe unserer Untersuchungen nicht verändert haben. Diese Annahme wird selbstverständlich in der Praxis nicht immer erfüllt sein. Verbesserungen in den Einrichtungen und der Organisation eines Betriebes können das Unfallrisiko vermindern, neue Arzttarife können den Erwartungswert der Kosten verschieben. Wir sprechen dann davon, dass infolge von äussern Entwicklungen sich der «wahre» Wert der Parameter geändert hat. Dieses Kapitel soll sich damit befassen, ein Kriterium zu entwickeln, auf Grund dessen wir entscheiden können, wenn solche Änderungen stattgefunden haben. Jedesmal, wenn wir dann auf Grund des Kriteriums eine solche Änderung annehmen müssen, wird es sich dann darum handeln, die Parameter neu zu schätzen.

Die Idee, welche dem in diesem Kapitel angewandten Verfahren zur Kontrolle der Parameter zugrunde liegt, kann mit dem Stichwort: «Hinreichende Statistik» (sufficient statistic) gekennzeichnet werden. Eine hinreichende Statistik für einen Parameter α ist folgendermassen definiert: Sie ist eine Funktion der Einzelbeobachtungen, so dass die bedingte mehrdimensionale Verteilungsfunktion der Einzelbeobachtungen bei gegebenem Wert dieser Funktion nicht mehr vom Parameter α abhängt. Sofern die Verteilungsfunktion der hinreichenden Statistik gewissen Bedingungen genügt (completeness), dann hängt diejenige Schätzung für α, welche minimale Varianz hat, nur von der hinreichenden Statistik ab. Da wir nun in der Regel keine unmittelbare Möglichkeit haben, einen Parameter zu beobachten, so helfen wir uns dadurch, dass wir die hinreichende Statistik für diesen Parameter kontrollieren. Wir werden dann entscheiden, dass der Parameter sich geändert hat, sobald die hinreichende Statistik eine signifikante Abweichung zeigt.

Im 4. Kapitel haben wir festgestellt, dass die Schätzungen der Parameter ν und δ^2 nur von den Beobachtungen (w_i, g_i) aus der Grundgesamtheit der Unfallhäufigkeiten abhängen, während die Schätzungen für γ und c allein durch die beobachteten Grössen (x_i, k_i) aus der Grundgesamtheit der Unfallkosten bestimmt werden. Es genügt deshalb, um Veränderungen der Parameter festzustellen, diese beiden Grundgesamtheiten getrennt zu kontrollieren.

a) Kontrolle der Grundgesamtheit der Unfallhäufigkeiten

Die mehrdimensionale Dichtefunktion für n Beobachtungstupel (w_i, g_i) hat folgende Form:

$$f(w_1, w_2, \ldots, w_n) = \frac{\prod_{i=1}^{n} \sqrt{g_i}}{(2\pi)^{\frac{n}{2}}\, \delta^n}\, e^{-\frac{1}{2\delta^2} \sum_{i=1}^{n} g_i (w_i - \nu)^2}$$

$$= \frac{\prod_{i=1}^{n} \sqrt{g_i}}{(2\pi)^{\frac{n}{2}}\, \delta^n}\, e^{-\frac{1}{2\delta^2} \left(\sum_{i=1}^{n} g_i w_i^2 - 2 \sum_{i=1}^{n} g_i w_i \nu + \sum_{i=1}^{n} g_i \nu^2 \right)}.$$

Aus dieser Form der Dichtefunktion erkennen wir unter Anwendung des Neymanschen «Faktorisierungstheorems»

α) $\sum g_i w_i^2$ ist eine hinreichende Statistik für δ^2,

β) $\sum g_i w_i$ ist eine hinreichende Statistik für $\frac{\nu}{\delta^2}$.

Um diese Grössen kontrollieren zu können, berechnen wir hier ihre Verteilungsfunktionen, auf Grund derer Kontrollkarten oder andere Testverfahren angewendet werden können.

a_α) $$w_i \approx N\left(\nu, \frac{\delta^2}{g_i}\right),$$

$$\frac{\sqrt{g_i}\, w_i}{\delta} \approx N\left(\frac{\sqrt{g_i}\, \nu}{\delta}, 1\right),$$

$\sum_{i=1}^{n} \frac{g_i w_i^2}{\delta^2}$ hat dann eine nichtzentrale χ^2-Verteilung mit n Freiheitsgraden,

$\sqrt{\sum \frac{g_i \nu^2}{\delta^2}}$ als Nichtzentralitätsparameter.

Symbolische Schreibweise

$$\sum_{i=1}^{n} \frac{g_i w_i^2}{\delta^2} \approx \chi_n^2\left(\sqrt{\sum \frac{g_i \nu^2}{2}}\right). \tag{20}$$

Bemerkung: Für numerische Berechnung siehe: Tables of noncentral χ^2, Evelyn Fix, University of California Publications in Statistics, 1949.

a_β)

$$w_i \approx N\left(\nu, \frac{\delta^2}{g_i}\right),$$

$$g_i w_i \approx N(g_i \nu, g_i \delta^2),$$

$$\sum_{i=1}^{n} g_i w_i \approx N\left(\sum_{i=1}^{n} g_i \nu, \sum_{i=1}^{n} g_i \delta^2\right). \tag{21}$$

b) Kontrolle der Grundgesamtheit der Unfallkosten

Die mehrdimensionale Dichtefunktion für n Beobachtungstupel (x_i, k_i) lautet:

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{(k_i c)^{k_i \gamma}}{\Gamma(k_i \gamma)} x_i^{k_i \gamma - 1} e^{-c k_i x_i}.$$

Das Neymansche «Faktorisierungstheorem» gibt uns in diesem Falle

α) $\sum_{i=1}^{n} k_i x_i$ ist hinreichend für c,

β) $\sum_{i=1}^{n} k_i \log x_i$ ist hinreichend für γ.

b_α)

$$x_i \approx \Gamma(k_i \gamma, k_i c),$$

$$k_i x_i \approx \Gamma(k_i \gamma, c),$$

$$\sum_{i=1}^{n} k_i x_i \approx \Gamma\left(\sum_{i=1}^{n} k_i \gamma, c\right), \tag{22}$$

b_β)

$$x_i \approx \Gamma(k_i \gamma, k_i c).$$

Die Dichtefunktion von x_i lautet:

$$f_{x_i}(x) = \frac{(k_i c)^{k_i\gamma}}{\Gamma(k_i\gamma)} x^{k_i\gamma-1} e^{-ck_i x},$$

diejenige von $\log x_i$ ist dann

$$f_{\log x_i}(u) = \frac{(k_i c)^{k_i\gamma}}{\Gamma(k_i\gamma)} e^{uk_i\gamma - k_i c e^u},$$

$$f_{k_i \log x_i}(t) = \frac{(k_i c)^{k_i\gamma}}{k_i\,\Gamma(k_i\gamma)} e^{t\gamma - k_i c e^{\frac{t}{k_i}}},$$

woraus sich durch Faltung die Dichte von $\sum k_i \log x_i$ ergibt:

$$f_{\Sigma k_i \log x_i}(t) = \prod_{i=1}^{n}{}^{*} f_{k_i \log x_i}(t). \tag{23}$$

Bemerkung: $\prod\limits_{i=1}^{n}{}^{*}$ soll symbolisch die Faltung der Dichten $f_{k_i \log x_i}(t)$ darstellen.

Ob diese Formel theoretischen Wert hat, sei dahingestellt. Auf jeden Fall ist sie aber für praktische Berechnungen unbrauchbar, und man wird sich deshalb nach geeigneten Approximationen umsehen. Wir rufen bei dieser Gelegenheit die Verwandtschaft der Verteilung mit der log. Normalverteilung in Erinnerung. Da der Verlauf der beiden Verteilungsfunktionen sehr ähnlich ist, wollen wir hier in Abänderung unserer Hypothese II annehmen, dass x_i logarithmisch normalverteilt ist.

Hypothese II:* $\log x_i \approx N(\mu_i, \sigma_i^2)$.

Um unsere modifizierte Hypothese mit der alten verträglich zu machen, berechnen wir die Parameter μ_i und σ_i^2 so, dass die zwei ersten Momente berechnet aus der log. Normalverteilung mit den aus der Γ-Verteilung bestimmten übereinstimmen.

Falls $\log x_i \approx N(\mu_i, \sigma_i^2)$, ergibt die Rechnung für Erwartungswert und Varianz von x_i:

$$E(x_i) = e^{\mu_i + \frac{\sigma_i^2}{2}}, \tag{24}$$

$$\mathrm{Var}\,(x_i) = e^{2\mu_i}\,(e^{2\sigma_i^2} - e^{\sigma_i^2}). \tag{25}$$

Andererseits legten wir in der (ursprünglichen) Hypothese II fest:

$$E(x_i) = \frac{\gamma}{c}, \tag{26}$$

$$\mathrm{Var}\,(x_i) = \frac{\gamma}{k_i c^2}. \tag{27}$$

Aus den Gleichungen (24)–(27) finden wir folgende Ausdrücke für die Parameter

$$\mu_i = \log\frac{\gamma}{c} - \frac{1}{2}\log\left(1 + \frac{1}{k_i\gamma}\right), \tag{28}$$

$$\sigma_i^2 = \log\left(1 + \frac{1}{k_i\gamma}\right). \tag{29}$$

Mittels dieser «Approximation» finden wir nun

$$\log x_i \approx N(\mu_i, \sigma_i^2),$$

$$k_i \log x_i \approx N(k_i\mu_i, k_i^2\sigma_i^2),$$

$$\sum_{i=1}^{n} k_i \log x_i \approx N\left(\sum_{i=1}^{n} k_i\mu_i; \sum_{i=1}^{n} k_i^2\sigma_i^2\right), \tag{30}$$

wobei μ_i und σ_i^2 aus den Formeln (28) und (29) zu substituieren sind.

Mit den Verteilungsfunktionen (20), (21), (22), (30) ist uns die Möglichkeit gegeben, auf die hinreichenden Statistiken die üblichen Testverfahren anzuwenden.

Literaturverzeichnis

R.S.Burington and D.C.May: Handbook of Statistics (Handbook Publishers Inc. Sandusky, Ohio, 1953).

M.Loève: Probability Theory (Van Nostrand, New York, 1955).

E.Fix: Tables of non central χ^2 (University of California Publications in Statistics, Vol.1, 1949).

K.Pearson: Tracts for Computers, Tables of the Digamma and Trigamma Functions (Cambridge University Press, 1919).

K.Pearson: Tables of the incomplete I-Function (Cambridge University Press, Re-issue, 1946).