Möglichkeiten der Anwendung von Anordnungstests im Versicherungswesen

Autor(en): **Berliner, B.**

Objekttyp: **Article**

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

Band (Jahr): **70 (1970)**

PDF erstellt am: **07.07.2024**

Persistenter Link: https://doi.org/10.5169/seals-967039

Möglichkeiten der Anwendung von Anordnungstests im Versicherungswesen

Von B. Berliner, Zürich

Einleitung

Die möglichen Resultate einer Statistik können als Punkte im Wahrscheinlichkeitsraum Ω aufgefasst werden. Die Wahrscheinlichkeit des ganzen Raumes, d.h. die Wahrscheinlichkeit, dass bei einer Beobachtung irgendein Ergebnis herauskommt, sei Eins. Eine Hypothese H_0 ordnet jedem messbaren Bereich $B \subset \Omega$ eine gewisse Wahrscheinlichkeit $PB = P(B/H_0)$ zu. Diese ist gleich dem Integral der Wahrscheinlichkeitsdichte über B. (Bei diskreten Wahrscheinlichkeitsverteilungen ist die Summation der Wahrscheinlichkeiten der einzelnen Punkte in B vorzunehmen.)

Ein Test wählt nun einen Bereich $V \subset \Omega$, den sogenannten Verwerfungsbereich, der eine Wahrscheinlichkeit von $P(V/\mathrm{H}_0) = z\%$ habe, aus, und schreibt vor, die Hypothese H_0 zu verwerfen, sobald der Beobachtungspunkt in den Verwerfungsbereich fällt.

Wenn die Hypothese H_0 richtig ist, dann ist unsere Irrtumswahrscheinlichkeit, d.h. die Wahrscheinlichkeit, dass wir die Hypothese H_0 zu Unrecht verwerfen, gleich $z\%$. Man nennt diese Irrtumswahrscheinlichkeit einen Fehler 1. Art.

Zu einer Hypothese H_0 können durch verschiedene Tests verschiedene Verwerfungsbereiche, die alle die gleiche Wahrscheinlichkeit von $z\%$ haben, gewählt werden. Wie soll man aus den überabzählbar vielen Tests bezüglich der Hypothese H_0 die geeignetsten herauswählen?

Es gibt zwei Kriterien, nach denen man sich richten soll, ein intuitives Kriterium und ein quantitatives Kriterium. Das intuitive Kriterium empfiehlt, einen Test abzulehnen, der einen Verwerfungsbereich definiert, von dem wir auf Grund des Gefühls oder der Erfahrung sagen würden, dass er ungeeignet ist. Die Testtheorie ist ein Verfahren, das nicht in der Lage ist, anzugeben, ob eine Hypothese mit einer Wahrscheinlichkeit von $z\%$ richtig ist, sondern nur bestimmt, wann eine Hypothese mit einer Irrtumswahrscheinlichkeit von $z\%$ zu verwerfen ist. Wir sind daher auf unsere Erfahrung bzw. auf unser Gefühl angewiesen, wenn wir beurteilen wollen, ob eine Beobachtung den Erwartungen entspricht oder aus dem Rahmen fällt. Einen Verwerfungsbereich können wir nur als geeignet bezeichnen, wenn er ausschliesslich aus dem Rahmen fallende Punkte im Wahrscheinlichkeitsraum enthält.

Wir können einen Teilaspekt dieses Sachverhalts noch etwas anders ausdrücken. Eine Verteilungsfunktion, d.h. eine Hypothese, wird so gewählt werden, dass sie einerseits die empirische, aus den Beobachtungen gewonnene, diskrete (Kolmogoroffsche) Verteilungsfunktion möglichst gut annähert und andererseits möglichst angenehme analytische Eigenschaften hat. Ein Test bezüglich einer solchen Hypothese ist dann geeignet, wenn er bei gegebener Irrtumswahrscheinlichkeit einen möglichst grossen Verwerfungsbereich definiert, d.h. einen Bereich, in den nur sehr seltene, von uns nicht erwartete Ereignisse fallen.

Das zweite, quantitative Kriterium für die Eignung eines Tests ist seine sogenannte Macht. Zu gegebenem Fehler 1. Art soll die Wahrscheinlichkeit eines Fehlers 2. Art, nämlich die Wahrscheinlichkeit des Nichtverwerfens einer falschen Hypothese, möglichst klein gemacht werden. Wir versuchen, die früher angegebene Schwäche der Testtheorie, dass sie nur negativer Aussagen fähig ist, also nur angeben kann, wann eine Hypothese zu verwerfen ist, aber nicht, ob sie richtig ist, dadurch zu mildern, dass wir gerade diejenigen Tests auswählen, die unsere Hypothese H_0 auszeichnen, indem sie sie möglichst verwerfen, wenn eine andere Alternativhypothese H_1 als richtig angenommen wird.

Sei H_0 eine Nullhypothese und $P(V/H_0) \leqq \frac{z}{100}$.

Nehmen wir nun eine Alternativhypothese H_1 als richtig an, dann ist der Fehler 2. Art $1-P(V/H_1)$ möglichst klein, also $M = P(V/H_1)$ möglichst gross zu machen. M heisst die Macht des Testes für H_0 in bezug auf die Alternative H_1.

Unter allen Tests $\varkappa$ mit Irrtumswahrscheinlichkeit $P(V_\varkappa/H_0) \leqq \frac{z}{100}$ gibt es in bezug auf H_1 einen mächtigsten. Dieser kann leicht konstruiert werden und heisst Likelihood Ratio Test. Die Lösung des Problems wird schwieriger, wenn H_1 eine zusammengesetzte Hypothese ist, die noch von Parametern (z.B. Mittelwert, Streuung) abhängt. Immerhin kann es unter Umständen einen bezüglich allen in H_1 enthaltenen Einzelhypothesen gleichmässig mächtigsten Test geben. Noch schwieriger, allgemeine Aussagen zu machen, wird es, wenn man auch noch zulässt, dass H_0 eine zusammengesetzte Hypothese ist. Die diesbezüglichen Theorien wurden von Neymann, Pearson, Scheffé und Lehmann entwickelt. Zum Schluss sei noch darauf hingewiesen, dass das intuitive und das quantitative Kriterium oft, durchaus aber nicht immer auf den gleichen Test führen.

Die Anordnungstests und die Möglichkeiten ihrer Anwendung im Versicherungswesen

Ein Grund, warum die Testtheorie bisher in der Versicherung allzusehr vernachlässigt wurde, ist ihr Unvermögen, positive Aussagen zu machen, mit andern Worten, eine Verteilungsfunktion mit einer bestimmten Wahrscheinlichkeit anzugeben. Sie ist daher ungeeignet für die Ausrechnung von Prämien, Streuungs- und Varianzzuschlägen usw.

Dagegen können Tests, insbesondere Anordnungstests, eine wichtige Kontrollfunktion ausüben, indem sie angeben, wann ein bestehendes System, z.B. ein Prämienstufensystem, aus dem Gleichgewicht gerät und revisionsbedürftig wird. Mit der Wahl der Grösse der Irrtumswahrscheinlichkeit hat man darüber hinaus ein quantitatives Instrument in der Hand. Je kleiner die Irrtumswahrscheinlichkeit, um so dringender und durchgreifender wird eine Neutarifierung nötig, falls ein Resultat in den Verwerfungsbereich fällt.

Mehrere Gesichtspunkte lassen die Anordnungstests als besonders geeignet für die Lösung der eben besprochenen Kontrollprobleme erscheinen.

1. An den beobachteten Grössen x, y zweier Zufallsvariabler X, Y (z.B. Grösse aller Männer und Frauen in der Schweiz) interessieren uns nicht die genauen Werte, sondern lediglich ihre Anordnung, also die Relationen $x < y$ und $x > y$ zwischen gemessenen x und y. Die Anordnungstests verwenden nur Anordnungen und

setzen keine bestimmte Verteilungsfunktion der Grössen X und Y voraus, sind also im Gegensatz etwa zu dem Chiquadrattest, dem Student-Test oder dem Varianzquotiententest verteilungsfrei. Dies sichert ihnen ein weitverbreitetes Anwendungsgebiet im Versicherungswesen zu.

2. Die Nullhypothese besagt, dass X und Y den gleichen Zentralwert haben oder, in einer schärferen Form, dass X und Y sogar die gleiche Verteilungsfunktion haben. Die Nullhypothese entspricht der Aussage, dass zwei bisher gleichtarifierte Risiken auch weiterhin gleich zu tarifieren sind. Die Anordnungstests prüfen, ob diese Aussage verworfen werden soll oder nicht, üben also genau die gewünschte Kontrollfunktion aus.
3. Zu der Nullhypothese existiert nur eine Alternativhypothese, die besagt, dass die Nullhypothese falsch ist. Es fallen somit alle Komplikationen, die bei zusammengesetzten Hypothesen auftreten, weg.
4. Es kommt oft vor, dass man die Daten eines kleinen Teilportefeuilles mit den Daten des Gesamtportefeuilles vergleichen will, um zu sehen, ob erstere Daten aus dem Rahmen fallen. Die meisten Anordnungstests besitzen den grossen Vorteil, auch bei kleinem statistischem Material zu gültigen Entscheidungen führen zu können.

Um Komplikationen zu vermeiden, nehmen wir in Zukunft immer an, die wahren Verteilungsfunktionen seien kontinuierlich und daher die Wahrscheinlichkeit $P(x_i = y_k) = 0$. Wir vermeiden so die Komplikationen mit den sogenannten Bindungen. Wenn wir einen Rappen als Geldeinheit annehmen, dann ist die Voraussetzung der Stetigkeit in ausgezeichneter Näherung erfüllt.

Die fünf im folgenden besprochenen Tests sollen nur kurz beschrieben und ihre Eigenschaften anhand von Beispielen illustriert werden. Vergleiche, welche Tests unter dem intuitiven und dem quantitativen Kriterium als am geeignetsten erscheinen, sollen angestellt werden.

Beispiel:

Das im folgenden aufgestellte Beispiel wird den ganzen Artikel hindurch verwendet werden, wobei es je nach Bedarf modifiziert werden wird.

Es seien auf Grund der Jahresstatistik einer Versicherungsgesellschaft aus 20 Ländern 1000 Schadenmeldungen in der Kaskoversicherung eingegangen. Der durchschnittlich geforderte Entschädigungsbetrag und der Zentralwert aus allen gemeldeten Schäden mögen beide sFr. 1000 betragen.

Aus einem gewissen europäischen Land seien im gleichen Zeitraum 10 Schadenmeldungen mit folgenden Beträgen eingegangen: Fr. 950, Fr. 1100, Fr. 1900, Fr. 2300, Fr. 2600, Fr. 2800, Fr. 3100, Fr. 3150, Fr. 3200, Fr. 10 000.

Gefühlsmässig würden wir meinen, die Ergebnisse dieses Landes würden, aus welchen Gründen auch immer, aus dem Rahmen fallen.

Wie gut bestätigen die einzelnen Tests diesen «Intuitiven Approach»?

Wir wollen mit einer Irrtumswahrscheinlichkeit von $\frac{z}{100} = \beta = 5\%$ arbeiten. Wir nehmen für alle folgenden Betrachtungen an, dass alle Grössen x_i, y_k unabhängig seien.

A. Der Zeichentest

Wir verwenden den einseitigen Test, da wir uns im Beispiel einseitig nur für die besonders schlechten Resultate interessieren. Wir bezeichnen den Zentral- bzw. Mittelwert von Fr. 1000 mit m. Die Nullhypothese besagt, dass für jeden Schaden j

$$P(x_j - m > 0) = P(x_j - m < 0) = \tfrac{1}{2} \quad \text{ist.}$$

Die Wahrscheinlichkeit, dass für das zu untersuchende Land von 10 Differenzen $x_i - m$, $i = 1, \ldots, 10$, 9 oder 10 grösser als Null sind, ist

$$\left[\binom{10}{9} + \binom{10}{10}\right] \cdot \left(\frac{1}{2}\right)^{10} = 11 \cdot \frac{1}{1024} = 0{,}01074,$$

also ungefähr 1%.

Da die Statistik Resultate aus 20 verschiedenen Ländern umfasst, sind wir auf Grund des Zeichentestes noch nicht berechtigt, dieses gewisse Land von der Versicherung zu den üblichen Tarifen auszuschliessen, denn schliesslich muss ja *eines* unter den 20 Ländern die «schlechteste» Schadenstatistik aufweisen. Wir haben aber gerade *dieses* Land gewählt, und die Wahrscheinlichkeit, dass unter 20 Ländern ein Land einen solch schlechten Schadenverlauf aufweisen wird, ist ungefähr $20 \cdot 0{,}01074 = 0{,}2148 \approx 21{,}5\%$, genau $1-(1-0{,}01074)^{20} = 0{,}194235 \approx 19{,}4\,\%$.

Wir haben das Gefühl, dass den *grossen Differenzen* zwischen den Schadenansprüchen des besprochenen Landes und dem durchschnittlichen Schadenanspruch nicht genügend Rechnung getragen wurde, denn schliesslich hätten wir auch das gleiche Resultat erhalten, wenn die 9 höheren Ansprüche zwischen Fr. 1000 und 1100 gelegen wären, in welchem Falle wir nie auch nur auf den Gedanken kämen, die Daten des entsprechenden Landes fielen aus dem Rahmen. Der Schadenanspruchsdurchschnitt des gewissen Landes X ist $m_x = \text{Fr. } 3110 > 3m$!

Der Grund für die unbefriedigende Antwort liegt in der Schwäche des Zeichentests. Um diesen anzuwenden, mussten wir eine Menge Informationen, die uns bekannt sind, unter den Tisch fegen. Viel gültigere Aussagen werden wir natürlich erhalten, wenn wir die 10 Schadenmeldungen ihrer Grösse nach mit *sämtlichen* übrigen 990 eingegangenen Schadenmeldungen vergleichen. Dies kann mit Hilfe des Wilcoxon-Tests oder des X-Tests erfolgen.

B. Der Wilcoxon-Test

Benennen wir die Schäden aus dem zu untersuchenden Land mit $x_i (i = 1, \ldots, g = 10)$ und die der übrigen Länder mit $y_k (k = 1, \ldots, h = 990)$. Die x_i und y_k werden nach steigender Grösse angeordnet. Eine Inversion ist der Austausch von einem x mit einem y, welches in der Folge unmittelbar vor dem betreffenden x steht. Es mögen so lange Inversionen vorgenommen werden, bis alle x am Anfang der Folge, d.h. links von allen y, stehen. Der einseitige Wilcoxon-Test schreibt nun vor, die Nullhypothese, dass alle möglichen Anordnungen gleich wahrscheinlich sind, zu verwerfen, sobald die Anzahl der Inversionen U eine Schranke U_β übersteigt.

Die Anzahl der möglichen Folgen ist

$$\binom{n}{g} = \frac{n!}{g!\,h!}, \quad \text{wobei } n = g + h\,.$$

In unserem Fall ist $\binom{n}{g} = \dfrac{1000!}{10!\,990!} \approx 2{,}6 \cdot 10^{23}$.

Wählen wir als Irrtumswahrscheinlichkeit 0,01%.

$$\beta = 0{,}0001 \qquad \beta\,\frac{1000!}{10!\,990!} \approx 2{,}6 \cdot 10^{19}$$

Wenn also weniger als $2{,}6 \cdot 10^{19}$ Situationen existieren, für die die Zahl der Inversionen grösser gleich der Anzahl der Inversionen im vorliegenden Fall ist, dann kann die Hypothese, dass die hohen Schäden aus dem Lande X rein zufällig sind, mit einer Wahrscheinlichkeit von ungefähr $1-20 \cdot 10^{-4} = 1-0{,}002 = 0{,}998 = 99{,}8\,\%$ verworfen werden. Wenn die Streuung in den Schadenansprüchen allgemein nicht *sehr* gross ist, dann wird der Wilcoxon-Test mit Sicherheit das oben angeführte signifikante Resultat liefern.

Statt der mühsamen oben erwähnten Methode können wir folgenden Satz verwenden:

Für festes g und $h \to \infty$ hat die Grösse

$$U = \sum z_{ik} \quad \text{mit} \quad z_{ik} = \begin{Bmatrix} 1 & \text{für } x_i > y_k \\ 0 & \text{sonst} \end{Bmatrix}$$

asymptotisch die Verteilungsfunktion einer Summe von g unabhängigen, zwischen 0 und h gleichverteilten Grössen. Ist ferner $g > 4$, dann kann bereits in ausgezeichneter Näherung folgender Satz verwendet werden:

Wenn g und h beide gegen ∞ gehen, ist U asymptotisch normal verteilt mit Mittelwert $\frac{1}{2}gh$ und Streuung

$$\sigma = \sqrt{\tfrac{1}{12}gh(g+h+1)}\,,$$

$$\mu = \tfrac{1}{2}\,990 \cdot 10 = 4950\,,$$

$$\sigma = (\tfrac{1}{12} \cdot 9900 \cdot 1001)^{\frac{1}{2}} = \sqrt{825825} = 908{,}75\,,$$

$$4\sigma = 3635 \qquad 3{,}7 \cdot \sigma = 3362\,,$$

$$\mu + 4\sigma = 8585 \quad \mu + 3{,}7 \cdot \sigma = 8312\,.$$

Übersteigt also die Anzahl der Inversionen die Zahl 8312, so ist die Hypothese mit einer Wahrscheinlichkeit von $1-20 \cdot 0{,}0001 = 0{,}998 = 99{,}8\%$ zu verwerfen.

C. Der X-Test

Sei gegeben die Testgrösse $X = \sum_{r_i} \Psi\left(\frac{r_i}{n+1}\right)$, wobei r_i die 10 Rangnummern der x_i sind, wenn alle $n = 1000$ Schadenansprüche in aufsteigender Grösse geordnet sind und Ψ die inverse Funktion der normalen Verteilungsfunktion Φ ist. Der einseitige X-Test schreibt vor, die Hypothese H_0, dass alle $n!$ Rangordnungen der x_i und y_k gleichwahrscheinlich seien, zu verwerfen, sobald X die Schranke X_β übersteigt.

Ein ähnliches Resultat wie beim Wilcoxon-Test ergibt sich beim X-Test durch folgenden Satz:

Seien alle Schadenansprüche x_i und y_k $(i = 1, \ldots, \quad g = 10)$ $(k = 1, \ldots, \quad h = 990)$ gleich verteilt und sei $h \gg g$. Dann ist X asymptotisch normal verteilt mit Mittelwert Null und Streuung

$$\sigma_x = \left[\frac{gh}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} \Psi^2\left(\frac{i}{n+1}\right)\right]^{\frac{1}{2}}.$$

Der X-Test ist so konstruiert, dass er für festes g und $h \to \infty$ im Fall der Normalverteilung genau so mächtig ist wie der Student-Test, der ja die grösstmögliche Macht hat. Der Wilcoxon-Test ist ebenfalls sehr mächtig. Seine asymptotische Effizienz ist $\frac{3}{\pi} \approx \frac{21}{22}$, mit andern Worten: Der Wilcoxon-Test, angewandt auf 22 Beobachtungen, ist ungefähr gleich mächtig wie der X-Test mit 21 Beobachtungen.

Um den Wilcoxon-Test einmal exakt durchzuexerzieren, wollen wir das frühere Beispiel dahingehend vereinfachen, dass wir nur noch eine Schadenstatistik aus zwei Ländern von je 10 Schadenfällen haben. Das erste Land weise Schadenfälle, wie sie auf Seite 353 angegeben sind, auf, während die Schäden aus dem zweiten Land durchwegs zwischen 950 und 1100 liegen mögen.

Der Ranggrösse nach geordnet, lauten dann die Schadenmeldungen:

$$x_1\, y_1\, y_2\, y_3\, y_4\, y_5\, y_6\, y_7\, y_8\, y_9\, y_{10}\, x_2\, x_3\, x_4\, x_5\, x_6\, x_7\, x_8\, x_9\, x_{10}\,.$$

x_i gehören zum ersten, y_k zum zweiten Land. Die Anzahl der Inversionen ist $9 \times 10 = 90$. Aus einer Tabelle für die Testwahrscheinlichkeit beim Wilcoxon-Test ergibt sich folgendes Resultat: Die Wahrscheinlichkeit, dass die Hypothese H_0 zu Unrecht verworfen wird, ist $2 \cdot 0{,}08\% = 0{,}16\,\%$. Der Faktor 2 kommt daher, dass wir jetzt den Test zweiseitig verwenden wollen, d.h. wir erlauben auch die Vertauschung der Rollen von den beiden Ländern.

Aus einer Tabelle für Schranken für X beim X-Test können wir folgendes ablesen: Ist für $n = g + h = 20$ und $g - h = 0$ $X > 4{,}94$, so ist die Hypothese H_0 (zweiseitig) mit einer Wahrscheinlichkeit von über 99% zu verwerfen.

$$X = \Psi\left(\frac{1}{21}\right) + \Psi\left(\frac{12}{21}\right) + \Psi\left(\frac{13}{21}\right) + \Psi\left(\frac{14}{21}\right) + \Psi\left(\frac{15}{21}\right) + \Psi\left(\frac{16}{21}\right) +$$

$$+ \Psi\left(\frac{17}{21}\right) + \Psi\left(\frac{18}{21}\right) + \Psi\left(\frac{19}{21}\right) + \Psi\left(\frac{20}{21}\right)$$

$$= \Psi(0{,}0476) + \Psi(0{,}5714) + \Psi(0{,}6190) + \Psi(0{,}6667) +$$

$$+ \Psi(0{,}7143) + \Psi(0{,}7619) + \Psi(0{,}8095) + \Psi(0{,}8571) +$$

$$+ \Psi(0{,}9048) + \Psi(0{,}9524) =$$

$$= -1{,}66 + 0{,}18 + 0{,}30 + 0{,}43 + 0{,}57 + 0{,}71 + 0{,}88 +$$

$$+ 1{,}07 + 1{,}31 + 1{,}66$$

$$= 5{,}45 > \underline{4{,}94}\,.$$

Im Falle des einseitigen Zeichentests erhalten wir genau wie auf Seite 353 eine Verwerfungswahrscheinlichkeit von 1,074%, wenn wir den Zentralwert m aus den y_k allein bilden.

Die Verwerfungsbereiche sind also beim Wilcoxon-Test und beim X-Test günstiger gewählt als beim Zeichentest. Da der Wilcoxon-Test und der X-Test mächtiger als der Zeichentest sind, sind sie ihm in unserem Beispiel sowohl wegen des intuitiven als auch wegen des quantitativen Kriteriums vorzuziehen.

Die Annahme, die x_i und die y_k wären gleich verteilt, wird vom Wilcoxon-Test viel entschiedener verworfen als vom X-Test. Wir geben daher ersterem Test den Vorzug, obwohl der X-Test, wenigstens asymptotisch, etwas mächtiger ist als der Wilcoxon-Test.

D. Der Smirnoff-Test

Sind die Schadenmeldungen $x_1, \ldots, x_g$ einerseits und die als Vergleichsbasis dienenden Schäden $y_1, \ldots, y_h$ andererseits gegeben und ist $k(t)$ als Anzahl der $x_i < t$ bzw. $l(t)$ als Anzahl der $y_k < t$ definiert, so erhalten wir die empirischen Verteilungsfunktionen

$$F_g(t) = \frac{k(t)}{g} \qquad \text{und} \qquad G_h(t) = \frac{l(t)}{h} \,.$$

Sei
$$\max_t / F_g(t) - G_h(t) / = D \,.$$

Ist $D > D_\beta$, so ist die Nullhypothese, die Verteilungsfunktionen von x und y seien gleich, zu verwerfen. D_β ist so bestimmt, dass $D > D_\beta$ eine Wahrscheinlichkeit β hat.

Eine gute Näherung für D_β für grosse n ist

$$D_\beta \approx \sqrt{-\tfrac{1}{2} \ln \beta} \left(\frac{1}{g} + \frac{1}{h} \right)^{\frac{1}{2}} .$$

Man bleibt mit dieser Formel immer auf der *sicheren Seite*. Der Smirnoff-Test hat gegenüber dem Wilcoxon-Test einige Nachteile. Er ist weniger mächtig. D_β hat die oben angegebene einfachere Form in guter Näherung nur für grosse n. Der Verwerfungsbereich, den der Smirnoff-Test auswählt, wird in den meisten im Versicherungswesen interessierenden Problemen weniger günstig sein als der Verwerfungsbereich unter dem Wilcoxon-Test. Dies wird im letzten Beispiel dieses Artikels deutlich werden. Will man aber aus bestimmten Gründen die völlige Übereinstimmung der Verteilungsfunktionen $F(t)$ und $G(t)$ der x und y in ihrem ganzen Verlauf prüfen, dann kann der Smirnoff-Test wegen seiner grossen Empfindlichkeit gegenüber jeder Abweichung zwischen den Verteilungsfunktionen mit grossen Vorteil angewendet werden.

Wenden wir diesen Test auf unser letztes Beispiel an:

$$D = / F_g(1099) - G_h(1099) / = 1 - \frac{1}{10} = \frac{9}{10} \,.$$

Sei $\beta = 0{,}01$

$$ln \frac{1}{100} = ln\, 100 = -2{,}3026 \cdot \log 100 = -4{,}6052$$

$$\sqrt{-\tfrac{1}{2}\, ln\, \beta} = \sqrt{2{,}3026} = 1{,}5174$$

$$D_\beta = 1{,}5174 \cdot \sqrt{0{,}2} = 1{,}5174 \cdot 0{,}4472 = 0{,}6786 < 0{,}9\,.$$

Sei $\beta = 0{,}001$

$$ln \frac{1}{1000} = -2{,}3026 \cdot 3$$

$$\sqrt{-\tfrac{1}{2}\, ln\, \beta} = \sqrt{3{,}4539} = 1{,}8585$$

$$D_\beta = 1{,}8585 \cdot 0{,}4472 = 0{,}8311 < 0{,}9\,!$$

Wir sehen hier die ausserordentliche Empfindlichkeit des Tests, der die Hypothese in unserem Beispiel mit einer Wahrscheinlichkeit, die grösser als 99,9% ist, verwerfen lässt.

Der Smirnoff-Test ist wegen der Verwendung von $D = / F - G /$ selbstverständlich ein zweiseitiger Test.

E. Der Kolmogoroff-Test

Mit Hilfe einer sehr umfangreichen Stichprobe habe man eine Verteilungsfunktion $F(t)$ für den Schadenverlauf bestimmt. Wie beim Smirnoff-Test wird nun für ein bestimmtes Land der empirische Schadenverlauf $F_g(t)$ definiert.

Die Hypothese, dass der Schadenverlauf in jenem bestimmten Land mit der Funktion $F(t)$ identisch sei, wird verworfen, wenn

$$\max_t / F(t) - F_g(t) / > \varepsilon\,,$$

wobei zu gegebenem β sich ε aus der Gleichung

$$\beta = 2 \sum_{i=1}^{\infty} (-1)^{i-1}\, 2^{-2i^2 n \varepsilon^2} \approx 2 e^{-2n\varepsilon^2} \qquad \text{bestimmt.}$$

Die funktionale Abhängigkeit zwischen β, ε und n ist tabelliert worden.

Der Kolmogoroff-Test ist dem Smirnoff-Test sehr nahe verwandt und weist daher auch dessen Vor- und Nachteile auf.

Ein letztes Beispiel:

Seien die y_k wie vorher $(950 < y_k < 1100;\ k = 1, \ldots, 10)$ und

$x_1 = \text{Fr. } 900$, $x_2 = \text{Fr. } 950$, $x_3 = \text{Fr. } 1100$, $x_4 = \text{Fr. } 1900$,

$x_5 = \text{Fr. } 2300$, $x_6 = \text{Fr. } 2600$, $x_7 = \text{Fr. } 2800$, $x_8 = \text{Fr. } 3100$,

$x_9 = \text{Fr. } 3150$, $x_{10} = \text{Fr. } 3200$

Wilcoxon-Test:

Anzahl der Inversionen $8 \times 10 = 80$.

Wahrscheinlichkeit, dass H_0 zu Unrecht verworfen wird: $2 \times 1{,}16 = 2{,}32\%$.

Legen wir eine Irrtumswahrscheinlichkeit von 5% zugrunde, so ist H_0 immer noch zu verwerfen.

X-Test:

$$\psi\left(\frac{1}{21}\right) + \psi\left(\frac{2}{21}\right) + \psi\left(\frac{14}{21}\right) + \ldots + \psi\left(\frac{20}{21}\right)$$

$$= +0{,}30 + 0{,}43 + 0{,}57 + 0{,}71 + 0{,}88 + 1{,}07$$

$$= \underline{3{,}96 > 3{,}86}\,.$$

Für $n = 20$; $g - h = 0$ ist bei einer zweiseitigen Irrtumswahrscheinlichkeit von 5% $X = 3{,}86$.

Zeichentest:

$$\left[\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right] \cdot \left(\frac{1}{2}\right)^{10} = (45 + 10 + 1)\,\frac{1}{1024} = \frac{56}{1024} =$$

$$= 0{,}0546 = 5{,}46\,\%\,.$$

Zeichen- und X-Test liefern hier sehr ähnliche Resultate.

Smirnoff-Test:

$$D = / F_g(1099) - G_h(1099) / = 1 - \tfrac{2}{10} = 0{,}8$$

$$\beta = 0{,}01: \qquad D_\beta = 1{,}5174 \cdot \sqrt{0{,}2} = 0{,}6786 < 0{,}8$$

$$\beta = 0{,}001: \qquad D_\beta = 0{,}8311 > 0{,}8\,.$$

Ist die Streuung der Schadenhöhen innerhalb einer Klasse sehr klein (wie in unserem Beispiel die Streuung der y_i), so kann der Smirnoff-Test sehr leicht zu einer ungerechtfertigten Verwerfung der Hypothese, dass die x_i und y_k im Durchschnitt gleich gross sind, führen. Wir haben hier im Gegensatz zum Zeichentest den Fall, dass ein Test ungeeignet wird, nicht weil er eine Hypothese nicht verwirft, wo wir es wünschen, sondern weil er eine Hypothese unter Umständen zu entschieden verwirft, also auch noch in solchen Fällen, wo wir eine Verwerfung aus gefühlsmässigen Gründen nicht mehr für gerechtfertigt halten würden.

Fazit: Der Wilcoxon-Test bzw. der X-Test können in vielen versicherungstechnischen Problemen, in denen die Berechtigung eines bestehenden Systems geprüft werden soll, mit Vorteil angewendet werden.

Schlussbemerkung

Wir haben gesehen, dass der Wilcoxon-Test, der X-Test und auch der Smirnoff-Test mehr Information aus den Daten verwenden, die uns zu Verfügung stehen, als der Zeichentest. Die Folge davon sind einschränkendere Nullhypothesen. Da in Versicherungsproblemen oft nur die Frage interessiert, ob die x und die y von gleicher Grössenordnung sind, haben wir bei der Beurteilung über die Eignung der Anordnungstests das intuitive Kriterium zu erweitern. Wir wählen den geeignetsten Test unter anderem nach dem Gesichtspunkt aus, wie gut der Verwerfungsbereich unserem Gefühl und unserer Erfahrung in bezug auf die *Alternativhypothese* der Gleichheit der Mittelwerte von X und Y entspricht. Da aus der Nullhypothese der Gleichheit der Verteilungsfunktionen die Alternativhypothese der Gleichheit der Mittelwerte folgt, müssen wir uns beim intuitiven Approach auf Eigenschaften der Nullhypothese konzentrieren, die mit der Alternativhypothese in Zusammenhang stehen. Andere Eigenschaften, wie z. B. die Abweichung der empirischen Verteilungsfunktionen voneinander in jedem Punkt sind unerheblich, da uns nur die Abweichung im Mittel interessiert. Ein Test wie der Smirnoff-Test, der einen Verwerfungsbereich nach Gesichtspunkten auswählt, die für uns uninteressant sind, wird also für unsere Zwecke ungeeignet sein.

Summary

Possibilities for applying five distribution-free tests in insurance with the help of so-called intuitive and quantitative criteria are examined. Tests – particularly distribution-free ones – can fulfill an important control function by indicating when an existing system loses its equilibrium and thus requires revision. In insurance this function is best performed by the Wilcoxon Test and the X-Test.

Zusammenfassung

Es werden die Anwendungsmöglichkeiten von fünf Anordnungstests im Versicherungswesen mit Hilfe sogenannter intuitiver und quantitativer Kriterien untersucht. Tests, und unter ihnen insbesondere die Anordnungstests, können eine wichtige Kontrollfunktion ausüben, indem sie angeben, wann ein bestehendes System aus dem Gleichgewicht gerät und revisionsbedürftig wird. Im Versicherungswesen werden dieser Aufgabe der Wilcoxon-Test und der X-Test am besten gerecht.

Résumé

L'auteur étudie les possibilités d'emploi de cinq tests non paramétriques, appliqués aux assurances, à l'aide de critères intuitifs et quantitatifs. Les tests, en particulier les tests non paramétriques, peuvent constituer un moyen de contrôle efficace, puisqu'ils permettent de déterminer quand un système établi n'est plus valable et doit être revisé. Appliqués aux assurances, les tests de Wilcoxon et de «X» sont les plus appropriés pour atteindre ce but.