ÜBER DAS ASYMPTOTISCHE VERHALTEN DER RAND- UND ZENTRALGLIEDER EINER VARIATIONSREIHE I

von

HANS-JOACHIM ROẞBERG[1]

Wir betrachten n unabhängige Zufallsgrössen $x_1, \ldots, x_n$ mit der Verteilungsfunktion $\mathbf{P}\{x_i < x\} = F(x)$. Bei einem Experiment an diesen x_i mögen die Werte X_i $(i = 1, \ldots, n)$ eingetreten sein. Dann ist die Ranggröße $\xi_k = R_k(X_1, \ldots, X_n)$ als der k-te Wert unter den X_i definiert, wenn diese der Größe nach angeordnet sind. Die so definierten Ranggrößen, die infolge der Beziehung $\xi_1 \leqq \xi_2 \leqq \ldots \leqq \xi_n$ voneinander abhängig sind, bilden eine sog. Variationsreihe. Unter den genannten Voraussetzungen hat die Theorie des Grenzverhaltens für $n \to \infty$ der Ranggrößen in vieler Hinsicht abschließende Resultate erzielt. Kaum behandelt wurden jedoch bisher die Beziehungen zwischen Rand- und Zentralgliedern einer Variationsreihe, die wir hier untersuchen wollen.

Wir beweisen einen Satz über die Unabhängigkeit im Limes von Ranggrößen. Was die Randglieder einer Variationsreihe betrifft, d. h. Ranggrößen ξ_h und ξ_k mit $h = \text{const}$ und $n - k = \text{const}$, so wurde ihre Unabhängigkeit im Limes (unter gewissen Voraussetzungen) von E. J. GUMBEL [2] festgestellt. Einen exakten Beweis dieser Tatsache und eine Verallgemeinerung lieferte T. HOMMA [3]. In einem sehr starken Sinne bewies J. GEFFROY [1] die Unabhängigkeit im Limes von Randgliedern. Den Fall zweier Zentralglieder $\left(\frac{h}{n} \to \lambda_1, \frac{k}{n} \to \lambda_2, \quad 0 < \lambda_1 < \lambda_2 < 1, \quad \text{d.h.} \quad \lim \frac{h}{k} > 0\right)$ behandelte N. W. SMIRNOFF [6]. Er zeigte, daß Unabhängigkeit im Limes nicht auftritt, wenn der Vektor (ξ_h, ξ_k) im Limes einer zweidimensionalen Normalverteilung unterworfen ist. Wir zeigen nun, daß Rand- und Zentralglieder stets im Limes unabhängig voneinander sind. In Bezug auf Formulierung und Beweis von Satz 1 verdanke ich Herrn Professor A. RÉNYI wertvolle Hinweise.

Satz 1. *Die Rangnummern h und k mögen derart von n abhängen, daß $\lim\limits_{n\to\infty} \frac{h}{k} = 0$. Dann sind die Ranggrößen ξ_h und ξ_k im Limes unabhängig, d. h. für*

$$\sigma_n = \sup_{\substack{-\infty<u<\infty \\ -\infty<v<\infty}} \left|\mathbf{P}(\xi_h < u, \xi_k < v) - \mathbf{P}(\xi_h < u)\,\mathbf{P}(\xi_k < v)\right|$$

gilt $\lim\limits_{n\to\infty} \sigma_n = 0$. In derselben Weise sind auch ξ_{n+1-h} und ξ_{n+1-k} im Limes voneinander unabhängig.

---

[1] Berlin.

Beweis. Die letzte Behauptung läßt sich auf die erste zurückführen, indem man die Zufallsgrößen $\bar{x}_i = -x_i$ heranzieht, die auf die Variationsreihe $\bar{\xi}_i = -\xi_{n+1-i}$ $(i = 1, \ldots, n)$ führen.

Wir definieren neue unabhängige Zufallsgrößen durch die Transformation

$$y_i = -\log(1 - F(x_i)) \qquad (i = 1, \ldots, n).$$

Sie geben Anlaß zu der Variationsreihe

$$\eta_i = -\log(1 - F(\xi_i)) \qquad (i = 1, \ldots, n).$$

Mit $\mathfrak{M}$ bezeichnen wir die Menge der Zahlen y, für die mindenstens eine Zahl $\hat{y}$ existiert, so daß $y = -\log(1 - F(\hat{y}))$. Für $y \in \mathfrak{M}$ gilt

$$\mathbf{P}\{y_i < y\} = G(y) = \mathbf{P}\{-\log(1 - F(x_i)) < -\log(1 - F(\hat{y}))\} = \\ = \mathbf{P}\{x_i < \hat{y}\} = F(\hat{y}) = 1 - e^{-y}.$$

Alle Zahlen $y \notin \mathfrak{M}$ liegen in Konstanzintervallen von $G(y)$. Wenn $F(x)$ stetig ist, haben somit die Zufallsgrößen y_i eine exponentielle Verteilungsfunktion. Im anderen Falle bildet der Wertevorrat von $G(y)$ nur eine echte Teilmenge des Intervalls $[0, 1]$. Nun ist

$$\sigma_n = \sup_{u,v} \big| \mathbf{P}\{\eta_h < -\log(1 - F(u)), \eta_k < -\log(1 - F(v))\} - \\ - \mathbf{P}\{\eta_h < -\log(1 - F(u))\} \mathbf{P}\{\eta_k < -\log(1 - F(v))\} \big| = \\ = \sup_{x,y \in \mathfrak{M}} \big| \mathbf{P}\{\eta_h < x, \eta_k < y\} - \mathbf{P}\{\eta_h < x\} \mathbf{P}\{\eta_k < y\} \big| .$$

Die Funktion, deren Supremum hier gebildet wird, läßt sich aber in der Form $Q_{hk}(G(x); G(y))$ darstellen. Die Verteilungsfunktion von η_k unter der Bedingung $\eta_h = u$ ist nämlich gleich der Verteilungsfunktion der Ranggröße mit der Nummer $k - h$ in einer Stichprobe vom Umfang $n - h$ mit der Grundverteilung

$$\frac{G(y) - G(u)}{1 - G(u)} \qquad (y \geqq u)$$

(vgl. [5], § 2). Daher gilt für $x \leqq y$

$$\mathbf{P}\{\eta_h < x, \eta_k < y\} = \int_0^x \mathbf{P}\{\eta_k < y \mid \eta_h = u\} \, d\mathbf{P}\{\eta_h < u\} = \\ = \binom{n}{h}\binom{n-h}{k-h} \int_0^{G(x)} \left\{ \int_0^{\frac{G(y)-w}{1-w}} (1 - s)^{n-k} \, ds^{k-h} \right\} d \int_0^w (1 - t)^{n-h} \, dt^h$$

und für $x > y$

$$\mathbf{P}(\eta_h < x, \eta_k < y) = \mathbf{P}\{\eta_k < y\} = \binom{n}{k} \int_0^{G(y)} (1 - t)^{n-k} \, dt^k .$$

Führen wir unabhängige Zufallsgrößen z_i mit $\mathbf{P}(z_i < z) = 1 - e^{-z}$ ($z \geqq 0$, $i = 1, \ldots, n$) ein, die die Variationsreihe $\zeta_1, \ldots, \zeta_n$ liefern, setzen wir ferner

$$\mathbf{P}\{\zeta_h < x,\ \zeta_k < y\} - \mathbf{P}\{\zeta_h < x\}\,\mathbf{P}\{\zeta_k < y\} = R_{hk}(x, y),$$

so gilt die Abschätzung

$$\sigma_n \leqq \sup_{x,y} |R(1 - e^{-x}, 1 - e^{-y})| = \sup_{x,y} |R_{hk}(x, y)|.$$

Damit ist unser allgemeines Problem auf ein viel einfacheres zurückgeführt, das wir mit Hilfe des folgenden Lemmas behandeln werden.

Lemma 1. *Wenn* $h \to \infty$, $\dfrac{h}{k} \to 0$, $n - k \to \infty$, *so ist gilt gleichmäßig bezüglich* x *und* y

$$R_{hk}(x, y) = \mathbf{P}\{\zeta_h < x,\ \zeta_k < y\} - \mathbf{P}\{\zeta_h < x\}\,\mathbf{P}\{\zeta_k < y\} \to 0.$$

Beweis. Da die Differenz $\zeta_k - \zeta_h$ unabhängig von ζ_h ist, (s. z. B. [4]), gilt

$$\begin{aligned} R_{hk}(x, y) &= \int_0^x [\mathbf{P}\{\zeta_k < y \mid \zeta_h = u\} - \mathbf{P}\{\zeta_k < y\}]\, d\mathbf{P}\{\zeta_h < u\} = \\ &= \int_0^x [\mathbf{P}\{\zeta_k - \zeta_h < y - u\} - \mathbf{P}\{\zeta_k < y\}]\, d\mathbf{P}\{\zeta_h < u\}. \end{aligned} \tag{1}$$

Man erkennt leicht, daß $R_{hk}(x, y) \geqq 0$ für alle x und y.

A. Rényi [4] hat eine Methode eingeführt, die es gestattet, Grenzwertsätze für Ranggrößen in einer sehr eleganten Weise zu gewinnen. Wir sind jetzt in der Lage, sie anzuwenden. Es gilt die Darstellung

$$\zeta_k = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \ldots + \frac{\delta_k}{n+1-k}, \tag{2}$$

wo die Zufallsgrößen $\delta_1, \ldots, \delta_k$ unabhängig sind und die Verteilungsfunktion $\mathbf{P}\{\delta_j < y\} = 1 - e^{-y}$ ($y \geqq 0$) haben, so daß für ihre mathematischen Erwartungen und Dispersionen

$$\mathbf{M}\,\delta_j = \mathbf{D}^2\,\delta_j = 1 \qquad (j = 1, \ldots, k)$$

gilt. Daraus folgt

$$\begin{aligned} M_k &= \mathbf{M}\,\zeta_k = \sum_{i=n+1-k}^{n} \frac{1}{i} = M_{kh} + M_h \\ S_k^2 &= \mathbf{D}^2\,\zeta_k = \sum_{i=n+1-k}^{n} \frac{1}{i^2} = S_{kh}^2 + S_h^2, \end{aligned} \tag{3}$$

und wegen $\mathbf{M}|\delta_j - 1|^3 = \int\limits_0^\infty |x-1|^3 e^{-x}\, dx < \int\limits_0^\infty e^{-x}\, dx + \int\limits_0^\infty x^3 e^{-x}\, dx \leqq 5$ gilt ferner

$$T_k^3 = \sum_{i=1}^{k} \mathbf{M} \left| \frac{\delta_i - 1}{n+1-i} \right|^3 \leqq 5 \sum_{i=n+1-k}^{n} \frac{1}{i^3} < 5 \int_{n-k}^{n} \frac{dt}{t^3} < \frac{5\,k}{n(n-k)^2}.$$

Durch derartige Integralabschätzungen (nach unten und oben) der Summen M_h und S_k^2 findet man weiter $(0 < \vartheta < 1)$

$$(4)\qquad \begin{aligned} M_h &= (1+o(1))\log\frac{n}{n-h} = \frac{h}{n}(1+o(1)) \\ S_k^2 &= \frac{k}{(n-k)\,n}\left(1+\vartheta\frac{n}{(n-k)\,k}\right) = \frac{k}{(n-k)\,n}(1+o(1)) \end{aligned}$$

und somit

$$\frac{T_k^3}{S_k^3} \leqq 5\sqrt{\frac{n}{(n-k)\,k}} \to 0\,.$$

Daher läßt sich auf die Summe (2) der Satz von Ljapunoff anwenden, d. h. es gilt

$$(5)\qquad \lim_{n\to\infty} \mathbf{P}\left\{\frac{\zeta_k - M_k}{S_k} < x\right\} = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt\,.$$

Wegen $M_{kh} = \mathbf{M}(\zeta_k - \zeta_h) = \sum\limits_{i=n+1-k}^{n-h}\frac{1}{i}$, $S_{kh}^2 = \mathbf{D}^2(\zeta_k - \zeta_h) = \sum\limits_{i=n+1-k}^{n-h}\frac{1}{i^2}$ braucht man in den obigen Rechnungen nur n und k durch $n-h$, bzw. $k-h$ zu ersetzen, um mit Hilfe von $\frac{h}{k}\to 0$ die Beziehungen

$$(6)\qquad S_{kh}^2 = S_k^2(1+o(1))$$

und

$$(7)\qquad \lim_{n\to\infty}\mathbf{P}\left\{\frac{\zeta_k - \zeta_h - M_{kh}}{S_{kh}} < x\right\} = \Phi(x)$$

zu gewinnen. Da die Konvergenz in (5) und (7) gleichmäßig bezüglich x ist, ergibt sich aus (1)

$$(8)\qquad \begin{aligned} R_{hk}(x,y) &= \int_0^x \left[\Phi\left(\frac{y-u-M_{kh}}{S_{kh}}\right) - \Phi\left(\frac{y-M_k}{S_k}\right)\right] d\,\mathbf{P}(\zeta_h < u) + r_n(x,y) = \\ &= J_n(x,y) + r_n(x,y) \end{aligned}$$

und das Restglied r_n konvergiert gleichmäßig in x und y gegen 0.

Bei beliebigem $y \geqq 0$ bestimmen wir die Zahl $\hat{x}(y)$ durch die Gleichung

$$\frac{y-\hat{x}-M_{kh}}{S_{kh}} = \frac{y-M_k}{S_k}, \quad \text{oder} \quad \hat{x} = \frac{y-M_k}{S_k}(S_k - S_{kh}) + M_h\,.$$

Nach (3) gilt $S_k - S_{kh} > 0$. Falls $\frac{y-M_k}{S_k} \to -\infty$ und dabei $\hat{x} < 0$, folgt wegen $\frac{y-M_{kh}}{S_{kh}} < \frac{y-M_k}{S_k}$ aus (8) die Behauptung des Lemmas. Wenn

$\frac{y-M_k}{S_k} > -C^2$, ist für genügend große n $\hat{x} > 0$, denn mit (4) und (6) folgt $\frac{S_k - S_{kh}}{M_h} = \frac{S_h^2}{M_h(S_k + S_{kh})} < k^{-\frac{1}{2}}$. Für $\hat{x} > 0$ aber ist bei festem y $\sup\limits_x J_n(x, y) = J_n(\hat{x}, y)$. Wenn $\frac{y - M_k}{S_k} \to \infty$, folgt nun aus (8) unmittelbar die Behauptung.

Es bleibt noch übrig, den Fall $\frac{y-M_k}{S_k} < C^2$, $\hat{x} > 0$ zu betrachten. Bei beliebigem $\eta > 0$ gilt die Abschätzung

$$J_n(\hat{x}, y) = \int\limits_0^{\hat{x}-\eta S_{kh}} \cdots + \int\limits_{\hat{x}-\eta S_{kh}}^{\hat{x}} \cdots \leqq \mathbf{P}\{\zeta_h < \hat{x} - \eta S_{kh}\} + \tag{9}$$

$$+ \left[\Phi\left(\frac{y-M_k}{S_k} + \eta\right) - \Phi\left(\frac{y - M_k}{S_k}\right)\right].$$

Das zweite Glied kann mit η gleichmäßig beliebig klein gemacht werden; das erste aber ist kleiner als $\mathbf{P}\left\{\frac{\zeta_h - M_h}{S_h} < C\,\frac{S_k - S_{kh}}{S_h} - \eta\,\frac{S_{kh}}{S_h}\right\}$ und verschwindet nach dem Satz von Ljapunoff für $n \to \infty$. Die Gleichung (5) gilt nämlich auch mit h an Stelle von k, und ebenso können wir in (4) k durch h ersetzen. Daher folgt mit Hilfe von (6) für genügend große n

$$\frac{S_k - S_{kh}}{S_h} = \frac{S_k^2 - S_{kh}^2}{S_h(S_k + S_{kh})} = \frac{S_h}{2S_k}(1 + o(1)) < \sqrt{\frac{h}{k}} \to 0$$

und

$$\frac{S_{kh}}{S_h} = \frac{S_k}{S_h}(1 + o(1)) \to \infty .$$

Damit ist das Lemma bewiesen.

Jetzt müssen wir uns noch von den Voraussetzungen $h \to \infty$, $n - k \to \infty$ befreien. Wir beginnen mit der letzteren, betrachten also die Funktion $R_{hl}(x, y)$ für $h \to \infty$, $\frac{h}{l} \to 0$, $n - l \not\to \infty$. Dann existiert eine Zahlenfolge $k(n)$, so daß $\frac{h}{k} \to 0$, $n - k \to \infty$, $k \leqq l$. Da die Ranggrößen $\zeta_1, \ldots, \zeta_n$ eine Markoffsche Kette bilden, sind ζ_h und ζ_l unter der Bedingung $\zeta_k = u$ unabhängig, und es gilt

$$R_{hl}(x, y) = \int\limits_0^y \mathbf{P}\{\zeta_l < x \mid \zeta_k = u\}\,[\mathbf{P}\{\zeta_h < x \mid \zeta_k = u\} - \mathbf{P}\{\zeta_h < x\}]\,d\mathbf{P}\{\zeta_k < u\} .$$

Natürlich brauchen wir nur Werte von y zu betrachten, für die der Integrand im Intervall $(0, y)$ positiv ist. Für solche Werte aber gilt

$$R_{hl}(x, y) \leqq \int_0^y [\mathbf{P}\{\zeta_h < x \mid \zeta_k = u\} - \mathbf{P}\{\zeta_h < x\}]\, d\mathbf{P}\{\zeta_k < u\} = R_{hk}(x, y)\,.$$

Auf Grund des Lemmas 1 ist damit der betrachtete Fall erledigt. Mit Hilfe dieses Resultats läßt sich aber in der gleichen Weise auch der Fall $h \nrightarrow \infty$ behandeln. Damit ist Satz 1 bewiesen.

(Eingegangen: 28. August 1960; in umgearbeiteter Form: 1. April 1962.)

LITERATURVERZEICHNIS

[1] Geffroy, J.: Publ. Inst. Stat. Paris 1958 und 1959.
[2] Gumbel, E. J.: Ann. Math. Stat. 1946.
[3] Homma, T.: Reports of statistical application research 1, Tokio, 1951.
[4] Rényi, A.: Acta Math. Acad. Sci. Hung. 1953.
[5] Roßberg, H. J.: Math. Nachrichten 1960.
[6] Smirnoff, N. W.: Bull. de l'Université d'État à Moscou, Ser. Inter., Section B, Math. et Méc. 1, 1937.

ОБ АСИМПТОТИЧЕСКОМ ПОВЕДЕНИИ ЦЕНТРАЛЬНОГО И КРАЙНЕГО ЧЛЕНОВ ПОРЯДКОВЫХ СТАТИСТИК

H. J. ROßBERG

Резюме

Используя метод A. Rényi автор доказывает, что упорядоченные статистики ξ_h и ξ_k являются асимптотически независимыми, если $n \to \infty$, $\frac{h}{k} \to 0$.