

# Bayes-Verfahren zur Schätzung von Einkommensverteilungen

Von Günter Bamberg

Eine Einkommensverteilung kann entweder durch eine stetige Verteilungskurve approximiert werden oder — bei Vorgabe von Einkommensgrößenklassen — in Histogrammform angegeben werden. Bei der Histogrammform ist die Verteilung durch die Klassenanteile festgelegt. Für die Schätzung dieser Klassenanteile mittels der Ergebnisse einer Stichprobenerhebung werden Formeln angegeben, die aus verschiedenen Bayes'schen Ansätzen resultieren.

## 1. Einleitung

Das Problem der Bayes-Schätzung einer Einkommensverteilung kann verschiedenartig angepackt werden:

a) Ein erster Weg besteht darin, daß zunächst eine ein- oder mehrparametrische Schar von stetigen Verteilungskurven als geeignete Approximationsmöglichkeit für die Einkommensverteilung ausgesondert wird, daß zusätzlich eine a priori Verteilung für die Parameter angegeben und schließlich Bayessche Punktschätzungen für die Parameter ermittelt werden. Parametrische Kurvenscharen, die in der Literatur für die Approximation von Einkommensverteilungen vorgeschlagen werden, sind die Klasse der Pareto-Verteilungen, der verschobenen logarithmischen Normal-Verteilungen, der Gamma-Verteilungen usw.; eine vergleichende Diskussion verschiedener Kurvenscharen findet man z. B. bei A. B. Z. Salem und T. D. Mount (1974).

Bei diesem ersten Weg taucht, überspitzt formuliert, folgendes Dilemma auf: Legt man eine relativ einfache, wenigparametrische Kurvenschar zugrunde, deren Parameter zudem einfach zu schätzen und anschaulich zu interpretieren sind (letzteres ist für die Angabe einer a priori Verteilung von Bedeutung), so ist die Anpassung an die Realität — zumindest in gewissen Einkommensbereichen — relativ schlecht. Legt man dagegen (zur Erhöhung der Flexibilität und zur Verbesserung der Anpassung) eine relativ komplexe, vielparametrische Schar zugrunde, so sind die Parameter in der Regel ziemlich unanschaulich, wodurch eine fundierte Angabe der a priori Verteilung unmöglich gemacht wird. Dieses Dilemma sowie die Häufung von Subjektivitäten (Auswahl der

Kurvenschar und Auswahl einer a priori Verteilung) wird bei dem folgenden Weg vermieden:

b) Die Einkommensverteilung wird nicht mehr als stetige Verteilungskurve, sondern als Histogramm geschätzt. Das Histogramm entsteht in natürlicher Weise durch die Bildung von Einkommensklassen. Nehmen wir an, daß  $k$  Einkommensklassen gebildet werden, so sind die  $k$  relativen Besetzungszahlen (= Besetzungswahrscheinlichkeiten bei einer Zufallsstichprobe)  $\vartheta_1, \dots, \vartheta_k$  zu schätzen. Damit hat man es zwar auch mit einer vielparametrischen (nämlich  $k$  bzw.  $(k - 1)$ -parametrischen) Kurvenschar zu tun; die Willkür bei der Auswahl der Kurvenschar ist jedoch verschwunden; außerdem besitzen alle Parameter eine höchst anschauliche Bedeutung.

## 2. Bayessche Histogrammschätzung bei Zugrundelegung einer speziellen a priori Verteilung und verschiedenen Schadensfunktionen

Es seien  $k$  (Einkommens-)Klassen gebildet mit den zugehörigen (unbekannten und zu schätzenden) Klassenanteilen  $\vartheta_1, \dots, \vartheta_k$ . Der Parameterraum ist demnach

$$\Theta = \left\{ \vartheta = (\vartheta_1, \dots, \vartheta_k) : \vartheta_1 \geq 0, \dots, \vartheta_k \geq 0, \sum_1^k \vartheta_i = 1 \right\}$$

Wird eine einfache Stichprobe vom Umfang  $n$  gezogen und werden die Besetzungszahlen für die  $k$  Klassen mit  $X_1, \dots, X_k$  bezeichnet<sup>1</sup>, so ist  $X = (X_1, \dots, X_k)$  multinomial-verteilt mit den Parametern  $n$  und  $\vartheta = (\vartheta_1, \dots, \vartheta_k)$ ; d. h. die Likelihoodfunktion auf dem Stichprobenraum

$$\mathfrak{X} = \left\{ (x_1, \dots, x_k) : x_i \geq 0, x_i \text{ ganzzahlig}, \sum_1^k x_i = n \right\}$$

ist

$$(1) \quad f(x | \vartheta) = \frac{n!}{x_1! \dots x_k!} \vartheta_1^{x_1} \dots \vartheta_k^{x_k}$$

<sup>1</sup> Da  $(X_1, \dots, X_k)$  eine suffiziente Statistik für  $(\vartheta_1, \dots, \vartheta_k)$  ist, können wir diese anstelle der ursprünglich gegebenen  $n$ -dimensionalen Stichprobenvariablen betrachten. Oft werden bei Stichprobenuntersuchungen ausschließlich diese Besetzungszahlen erhoben; so sind beispielsweise in den Erhebungsbögen für den Mikrozensus (seit 1972) 15 Einkommensgrößenklassen vorgegeben. Dies hat seinen Grund in der Tatsache, daß die Interviewten erstens ihr Einkommen vielfach nicht exakt kennen und zweitens „wesentlich leichter dazu zu bewegen sind, sich in vorgegebene, grob gegliederte Einkommensgrößenklassen einzustufen, als sämtliche Einnahmen auf Heller und Pfennig nachzuweisen.“ (M. Euler 1974, S. 74).

Wählen wir als a priori Verteilung  $\varphi$  auf  $\Theta$  eine allgemeine Beta-Verteilung, auch *Dirichlet*-Verteilung genannt, mit dem positiven Parametervektor  $(\mu_1, \dots, \mu_k)$ , so ist

$$(2) \quad \varphi(\vartheta) = \frac{\Gamma(\mu_1 + \dots + \mu_k)}{\Gamma(\mu_1) \dots \Gamma(\mu_k)} \vartheta_1^{\mu_1-1} \dots \vartheta_k^{\mu_k-1}$$

zu setzen. Dabei besteht zwischen den Parametern  $\mu_1, \dots, \mu_k$  sowie den Erwartungswerten und Varianzen der  $\vartheta_i$  folgender Zusammenhang (vgl. *M. H. De Groot* 1970, S. 51):

$$(3) \quad E(\vartheta_i) = \frac{\mu_i}{\sum \mu_j}, \text{Var}(\vartheta_i) = \frac{\mu_i(\sum \mu_j - \mu_i)}{(\sum \mu_j)^2(1 + \sum \mu_j)} ;$$

insbesondere entsprechen die Verhältnisse zwischen den  $\mu_i$  den Verhältnissen zwischen den Erwartungswerten  $E(\vartheta_i)$ , so daß die a priori Verteilung (2) im wesentlichen bereits durch die Schätzung der a priori Erwartungswerte  $E(\vartheta_i)$  bestimmt ist.

Für die a posteriori Verteilung  $\psi$  ergibt sich durch Produktbildung von (1) und (2) die Proportionalität

$$\psi(\vartheta | x) \sim \vartheta_1^{x_1+\mu_1-1} \dots \vartheta_k^{x_k+\mu_k-1}$$

Die a posteriori Verteilung von  $\vartheta$  ist also eine *Dirichlet*-Verteilung mit dem Parametervektor  $(x_1 + \mu_1, \dots, x_k + \mu_k)$ . Die Marginalverteilung von  $\vartheta_i$  ist a priori eine Beta-Verteilung  $\varphi_i(\vartheta_i)$  mit dem Parametervektor  $(\mu_i, \sum \mu_j - \mu_i)$  und a posteriori eine Beta-Verteilung  $\psi_i(\vartheta_i | x)$  mit dem Parametervektor  $(x_i + \mu_i, n + \sum \mu_j - x_i - \mu_i)$ .

Bezeichnen wir — wie in der Schätztheorie üblich — den Schätzwert für  $\vartheta$  bzw.  $\vartheta_i$  mit  $\hat{\vartheta}$  bzw.  $\hat{\vartheta}_i$ , so ergibt sich:

a) Bei Zugrundelegung der quadratischen Schadensfunktion

$$s(\vartheta, \hat{\vartheta}) = \|\vartheta - \hat{\vartheta}\|^2 = \sum_1^k (\vartheta_i - \hat{\vartheta}_i)^2$$

ist der a posteriori Schadenserwartungswert

$$\int \sum_1^k (\vartheta_i - \hat{\vartheta}_i)^2 \psi(\vartheta | x) d\vartheta = \sum_1^k \int_0^1 (\vartheta_i - \hat{\vartheta}_i)^2 \psi_i(\vartheta_i | x) d\vartheta_i .$$

zu minimieren. Bekanntlich<sup>2</sup> ist  $\hat{\vartheta}_i$  gleich dem a posteriori Erwartungswert, in unserem Falle also

<sup>2</sup> Vgl. z. B. *G. Bamberg* (1972), S. 104.

$$(4) \quad \hat{\vartheta}_i = \frac{x_i + \mu_i}{n + \sum \mu_j}$$

zu setzen.

b) Bei Zugrundelegung der Betrags-Schadensfunktion

$$s(\vartheta, \hat{\vartheta}) = \sum_1^k |\vartheta_i - \hat{\vartheta}_i|$$

muß  $\hat{\vartheta}_i$  gleich dem Median der a posteriori Verteilung  $\psi_i$  gesetzt werden (vgl. z. B. H. Raiffa, R. Schlaifer 1961, S. 196). Für die numerische Auswertung sind Tabellen erforderlich (z. B. K. Pearson 1934).

c) Bei Zugrundelegung der Null-Eins-Schadensfunktion mit Indifferenzbereich  $\pm \varepsilon$

$$s(\vartheta, \hat{\vartheta}) = \begin{cases} 0, & \text{für } |\vartheta - \hat{\vartheta}| \leq \varepsilon \\ 1, & \text{für } |\vartheta - \hat{\vartheta}| > \varepsilon \end{cases}$$

muß  $\hat{\vartheta}$  gleich dem Modalwert von  $\psi_i$ , also

$$(5) \quad \hat{\vartheta}_i = \frac{x_i + \mu_i - 1}{n + \sum \mu_j - k}$$

gesetzt werden<sup>3</sup>.

Bemerkungen zu den Bayes-Schätzern (4) und (5):

1. Der für festes  $\vartheta$  gebildete Erwartungswert von (4) ist

$$E(\hat{\vartheta}_i) = \vartheta_i \frac{n}{n + \sum \mu_j} + \frac{\mu_j}{n + \sum \mu_j} \neq \vartheta_i ;$$

der entsprechende Erwartungswert von (5) ist

$$E(\hat{\vartheta}_i) = \vartheta_i \frac{n}{n - k + \sum \mu_j} + \frac{\mu_j - 1}{n - k + \sum \mu_j} \neq \vartheta_i .$$

Die Bayes-Schätzer (4) und (5) sind — im Gegensatz zum Maximum-Likelihood-Schätzer  $\tilde{\vartheta}_i = \frac{x_i}{n}$  — nicht erwartungstreu.

2. Die Verzerrtheit der Schätzfunktionen — die man in der Regel auch bei anderen Bayes-Schätzfunktionen vorfindet, so auch im nächsten Abschnitt bei der Schätzfunktion (12) bzw. (14) und (15) — liegt darin

<sup>3</sup> Hierbei ist vorausgesetzt — was in der Praxis stets erfüllt sein dürfte, falls die Klassen nicht zu klein gewählt werden —, daß  $x_i + \mu_i \geq 1$  ist ( $i = 1, \dots, k$ ).

begründet, daß man nicht nur die „Daten sprechen läßt“, sondern den aufgrund der reinen Stichprobenrealisation nahegelegten Schätzwert in Richtung auf den Wert „zerrt“, der durch die a priori Verteilung nahegelegt wird. Diese a prioristische Voreingenommenheit für bestimmte Werte mag vom wissenschaftstheoretischen Standpunkt aus verdächtig erscheinen und die Verwendung von Bayes-Methoden bei solchen Fragestellungen verbieten, bei denen es ausschließlich um reine Erkenntnisziele geht. Die Bayes-Methode ist aber durchaus angebracht, wenn pragmatischere, entscheidungsorientierte Fragestellungen im Vordergrund stehen, wie z. B.:

- Wie kann man bei gegebenem Budget für die empirische Untersuchung (bzw. gegebenem Stichprobenumfang) zu einer möglichst guten Entscheidung gelangen?
- Wie geht man angesichts bestimmter Stichproben- und Fehlentscheidungskosten am besten vor, d. h. wie ist der Stichprobenumfang zu wählen und die auf diesem Stichprobenumfang basierende Stichprobenrealisation auszuwerten?

3. Legt man als a priori Verteilung eine symmetrische *Dirichlet*-Verteilung zugrunde, d. h. setzt man  $\mu_1 = \mu_2 = \dots = \mu_k = \mu$ , so wird aus (4) die Schätzfunktion

$$\hat{\vartheta}_i = \frac{x_i + \mu}{n + k \mu},$$

die von *I. J. Good* (1965, S. 24) als „glättende Schätzfunktion“ bezeichnet wird<sup>4</sup>.

Setzt man zusätzlich  $\mu = 1$  (Gleichverteilung auf  $\Theta$ ), so wird aus (5) die Maximum-Likelihood-Schätzfunktion und aus (4) die Schätzfunktion

$$\hat{\vartheta}_i = \frac{x_i + 1}{n + k},$$

die bereits von *G. J. Lidstone* (1920) angegeben wurde und für  $k = 2$  in die aus dem „Laplaceschen Nachfolgesatz“ bekannte Formel übergeht.

4. Normiert man die Parameter  $\mu_i$  durch  $\sum \mu_i = 1$ , so ist  $\mu_i$  nach (2) mit  $E(\vartheta_i)$  identisch; die Schätzfunktionen (4) bzw. (5) unterscheiden sich wegen der Kleinheit von  $\mu_i$  nur unwesentlich von der Maximum-Likelihood-Schätzfunktion. Multipliziert man alle  $\mu_i$  beispielsweise mit dem Faktor 1000, so bleiben die a priori Erwartungswerte  $E(\vartheta_i)$  zwar dieselben; der in (4) bzw. (5) enthaltene „Bayessche Korrektoreinfluß“ ist

---

<sup>4</sup> Weitere Bayessche Glättungsmethoden für Histogramme wurden von *P. Whittle* (1957, 1958), *J. M. Dickey* (1968) und *T. Leonard* (1973) behandelt.

nun aber beträchtlich. Dies liegt daran, daß im letzteren Fall die a priori Varianz  $\text{Var}(\vartheta_i)$  weitaus kleiner als im ersten Fall ist (vgl. (2)) und mithin der a priori Verteilung, insbesondere dem a priori Erwartungswert  $E(\vartheta_i)$ , ein erhöhtes Gewicht beigemessen wird. Man benötigt also für (4) bzw. (5) neben der Schätzung von  $E(\vartheta_i)$  auch eine Schätzung für  $\text{Var}(\vartheta_i)$ . Zweckmäßigerweise geht man hierbei zweistufig vor; betrachten wir hierzu folgendes numerische Beispiel:

Die a priori Erwartungswerte liefern normierte Parameter  $\mu_i$ , von denen etwa  $\mu_1 = 0,2$  der größte sei. Dann überlegt man sich, ob die a priori Varianz von  $\vartheta_1$

$$(6) \quad \frac{\mu_1(1 - \mu_1)}{2} = \frac{4}{100}$$

adäquat ist oder kleiner ausfallen könnte. Die Varianz (6) ist im Verhältnis zum Erwartungswert  $\mu_1 = 0,2$  offensichtlich riesig; der Zwei-Sigma-Bereich beträgt  $0,2 \pm 2 \cdot 0,2$ , also  $20\% \pm 40\%$ . Sei etwa eine a priori Varianz von  $\left(\frac{1}{40}\right)^2$  angemessen; der Zwei-Sigma-Bereich beträgt dann immerhin noch  $20\% \pm 5\%$ . Diese Varianz ergibt sich, wenn alle  $\mu_i$  mit einem Faktor  $\alpha$  multipliziert werden, der sich nach (2) aus

$$\frac{\alpha \cdot 0,2(\alpha - 0,2\alpha)}{\alpha^2(1 + \alpha)} = \left(\frac{1}{40}\right)^2$$

zu  $\alpha = 255$  berechnet. In die Formeln (4) bzw. (5) muß für  $\mu_i$  also das 255fache des a priori Erwartungswertes  $E(\vartheta_i)$  eingesetzt werden.

### 3. Lineare Bayessche Histogrammschätzung

Da es nur wenige „handliche“ Verteilungen auf  $\Theta$  gibt und die rechen-technisch einfachste — die *Dirichlet*-Verteilung — auf Schätzfunktionen (4) bzw. (5) führt, die in  $x_i$  linear sind, liegt der Versuch nahe, auf die Festlegung einer speziellen a priori Verteilung zu verzichten und ausschließlich die (in  $x_i$ ) linearen Schätzfunktionen zu untersuchen.

Wir beschränken uns also auf die Klasse  $\mathcal{A}$  der Schätzfunktionen des Typs

$$(7) \quad \hat{\vartheta}_i = \alpha_i + \beta_i x_i \quad (i = 1, \dots, k)$$

mit Konstanten  $\alpha_i, \beta_i$ , die von  $x_i$  unabhängig sind.

Die Nebenbedingung

$$(8) \quad \sum_1^k \hat{\vartheta}_i = \sum_1^k \alpha_i + \sum_1^k \beta_i x_i = 1$$

muß insbesondere für  $(x_1, \dots, x_k) = (n, 0, \dots, 0), (0, n, 0, \dots, 0), \dots$  und  $(0, \dots, 0, n)$  gelten und liefert

$$\beta_1 = \dots = \beta_k = \frac{1}{n} \left( 1 - \sum_1^k \alpha_i \right).$$

Wir können uns deshalb auf die Betrachtung von Schätzfunktionen des Typs

$$(9) \quad \hat{\vartheta}_i = \alpha_i + \beta x_i \quad (i = 1, \dots, k)$$

mit

$$(10) \quad \sum \alpha_i + n \beta = 1$$

beschränken. Legen wir zusätzlich die quadratische Schadensfunktion zugrunde, so ergibt sich bzgl. der a priori Verteilung  $\varphi(\vartheta)$  der Risikoerwartungswert

$$(11) \quad \begin{aligned} r(\varphi, \vartheta) &= \sum_{i=1}^k E_\varphi E_\vartheta (\alpha_i + \beta X_i - \vartheta_i)^2 = \\ &= \sum_{i=1}^k E_\varphi E_\vartheta (\alpha_i^2 + 2 \alpha_i \beta X_i + \beta^2 X_i^2 - 2 \alpha_i \vartheta_i - 2 \beta \vartheta_i X_i + \vartheta_i^2) \\ &= \sum_{i=1}^k [\alpha_i^2 + 2 \alpha_i \beta n E(\vartheta_i) + \beta^2 E(X_i^2) - 2 \alpha_i E(\vartheta_i) - 2 \beta n E(\vartheta_i^2) + \\ &\quad + E(\vartheta_i^2)] \end{aligned}$$

Die Minimierung des Risikoerwartungswertes (11) unter der Nebenbedingung (10) liefert (vgl. Anhang, 4.1)

$$\alpha_i = E(\vartheta_i) \left[ 1 - n^2 \frac{\sum \text{Var}(\vartheta_i)}{\sum \text{Var}(X_i)} \right]$$

und

$$\beta = n \frac{\sum \text{Var}(\vartheta_i)}{\sum \text{Var}(X_i)},$$

und damit als optimale lineare Schätzfunktion:

$$(12) \quad \hat{\vartheta}_i = E(\vartheta_i) + [x_i - n E(\vartheta_i)] n \frac{\sum_1^k \text{Var}(\vartheta_i)}{\sum_1^k \text{Var}(X_i)}$$

Bemerkungen zur Bayes-Schätzfunktion (12):

1. Benutzt man die Abkürzungen

$$\mu_i = E(\vartheta_i), \quad \sigma_i^2 = \text{Var}(\vartheta_i)$$

für den a priori Erwartungswert bzw. die a priori Varianz von  $\vartheta_i$ , und drückt man  $\Sigma \text{Var}(X_i)$  gemäß

$$(13) \quad \sum_1^k \text{Var}(X_i) = n - n \Sigma \mu_i^2 + n(n-1) \Sigma \sigma_i^2$$

durch  $n$ ,  $\mu_i$  und  $\sigma_i^2$  aus (vgl. Anhang, 4.2), so nimmt (12) folgendes Aussehen an:

$$(14) \quad \hat{\vartheta}_i = \mu_i + (x_i - n \mu_i) \frac{\Sigma \sigma_i^2}{1 - \Sigma \mu_i^2 + (n-1) \Sigma \sigma_i^2}$$

Man ersieht aus (14), daß neben der Angabe der a priori Erwartungswerte  $\mu_i$  nur die Angabe der Summe aller a priori Varianzen  $\sigma_i^2$ , nicht dagegen die Angabe der einzelnen  $\sigma_i^2$  selbst oder weiterer a priori Verteilungsannahmen erforderlich ist.

2. Die Schätzwerte (12) bzw. (14) liegen stets im Intervall  $[0,1]$  (vgl. Anhang, 4.3).

3. Setzt man in (14) die Varianzsumme  $\Sigma \sigma_i^2$  Null, so liefert die Schätzfunktion stets den a priori Erwartungswert  $\mu_i$  als Schätzwert; die in diesem Fall gegebene dogmatische Voreingenommenheit für  $\mu_i$ , die in der einpunktigen a priori Verteilung enthalten ist, spiegelt sich auch in der Schätzfunktion wieder. Der Schätzwert ist ferner dann mit  $\mu_i$  identisch, wenn die Differenz

$$x_i - n \mu_i$$

aus der empirisch beobachteten und der a priori zu erwartenden Besetzungszahl verschwindet.

Mit wachsendem Stichprobenumfang  $n$  (und positiver Varianzsumme) konvergiert  $\hat{\vartheta}_i$  gegen die Maximum-Likelihood-Schätzfunktion  $\frac{x_i}{n}$ .

4. Die Schätzfunktion  $\hat{\vartheta}_i$  läßt sich auch als gewichtetes Mittel aus dem Likelihoodschätzwert  $\frac{x_i}{n}$  und dem a priori zu erwartenden Anteilswert  $\mu_i$  darstellen<sup>5</sup>:

$$(15) \quad \hat{\vartheta}_i = \lambda \frac{x_i}{n} + (1 - \lambda) \mu_i,$$

wobei

<sup>5</sup> Eine ähnliche Darstellung errechnet *M. Goldstein* (1975) für die Bayes-Schätzung einer Verteilungsfunktion.



$$\lambda = \frac{\sum \sigma_i^2}{\frac{1 - \sum \mu_i^2}{n} + \frac{n - 1}{n} \sum \sigma_i^2}$$

gesetzt wurde.

Der Gewichtungsfaktor  $\lambda$  ist bei großem Stichprobenumfang  $n$  und relativ unpräziser a priori Information, d. h. großem  $\sigma_i$ , nahe bei 1. Bei stark disaggregiertem Untersuchungsziel, etwa bei der Schätzung der Einkommensverteilung der 30- bis 40jährigen Diplomingenieure, dürfte der Stichprobenumfang dagegen in der Regel so klein sein, daß auch  $\lambda$  relativ weit unter 1 liegt. Ist beispielsweise  $n = 500$ ,  $k = 5$ ,  $\sigma_i = 0,02$  (der  $3 \sigma$ -Bereich also  $\pm 6 \%$ ) und ist  $\mu_1 = \dots = \mu_5 = 0,2$ , so beträgt  $\lambda = 0,56$ . Der lineare Bayes' Schätzer führt hier also annähernd zu einer fifty-fifty-Gewichtung aus der a priori- und der Stichprobeninformation.

5. Geht es lediglich um die Schätzung eines einzigen Anteilswertes  $\vartheta$  — etwa um den Anteil derjenigen Haushalte mit einem Jahreseinkommen über DM 100 000 — so ist in (12) speziell  $k = 2$  sowie

$$\begin{aligned} \vartheta_1 &= \vartheta & (\vartheta_2 &= 1 - \vartheta_1) \\ \mu_1 &= \mu, \mathbf{X}_1 = \mathbf{X}, x_1 = x \end{aligned}$$

zu setzen. Dann erhält man die lineare Bayessche Schätzfunktion

$$\hat{\vartheta} = \mu + [x - n \mu] \frac{n \text{Var}(\vartheta)}{\text{Var}(\mathbf{X})}$$

für den Parameter  $\vartheta$  einer Binomialverteilung; diese Schätzfunktion findet sich in einer Arbeit von *B. S. Griffin* und *R. G. Krutchkoff* (1971) und wurde dort zu einer empirischen Bayesschen Schätzfunktion ausgebaut.

#### 4. Anhang

##### 4.1 Minimierung von (11) unter der Nebenbedingung (10)

Die Lagrange-Funktion mit dem Lagrange-Parameter  $2 \lambda$  lautet:

$$\begin{aligned} \Phi(\alpha_i, \beta, \lambda) &= \sum_1^k [\alpha_i^2 + 2 \alpha_i \beta n E(\vartheta_i) + \beta^2 E(\mathbf{X}_i^2) - 2 \alpha_i E(\vartheta_i) - \\ &\quad - 2 \beta n E(\vartheta_i^2) + E(\vartheta_i^2)] - 2 \lambda [\sum \alpha_i + n \beta - 1] . \end{aligned}$$

Die notwendigen Bedingungen sind

$$(16) \quad \frac{\partial \Phi}{\partial \alpha_i} = 2 \alpha_i + 2 \beta n E(\vartheta_i) - 2 E(\vartheta_i) - 2 \lambda = 0$$

$$(17) \quad \frac{\partial \Phi}{\partial \beta} = \sum_1^k [2 \alpha_i n E(\vartheta_i) + 2 \beta E(X_i^2) - 2 n E(\vartheta_i^2)] - 2 \lambda n = 0$$

Aus (16) ergibt sich

$$(18) \quad \alpha_i = \lambda + E(\vartheta_i) - \beta n E(\vartheta_i) ,$$

und aus (17) ergibt sich

$$(19) \quad \beta \Sigma E(X_i^2) + n \Sigma \alpha_i E(\vartheta_i) = n \Sigma E(\vartheta_i^2) + \lambda n .$$

Setzt man (18) in (19) ein, so erhält man sukzessive

$$\beta \Sigma E(X_i^2) + n \Sigma E(\vartheta_i) [\lambda + E(\vartheta_i) - \beta n E(\vartheta_i)] = n \Sigma E(\vartheta_i^2) + \lambda n$$

$$\beta \Sigma E(X_i^2) + n \Sigma E^2(\vartheta_i) - \beta \Sigma n^2 E^2(\vartheta_i) = n \Sigma E(\vartheta_i^2)$$

$$\beta [\Sigma E(X_i^2) - \Sigma E^2(X_i)] = n \Sigma [E(\vartheta_i^2) - E^2(\vartheta_i)]$$

$$\beta \Sigma \text{Var}(X_i) = n \Sigma \text{Var}(\vartheta_i)$$

$$\beta = n \frac{\Sigma \text{Var}(\vartheta_i)}{\Sigma \text{Var}(X_i)} .$$

Setzt man dies in (18) ein und nutzt man (10) aus, so ergibt sich  $\lambda = 0$  und  $\alpha_i$  wie in Abschnitt 3.

Die hinreichenden Bedingungen für ein Minimum sind ebenfalls erfüllt, denn der Risikoerwartungswert (11) ist eine streng konvexe Funktion von  $\alpha_1, \dots, \alpha_k, \beta$  (die im Unendlichen kein Minimum besitzen kann). Die strenge Konvexität von (11) ergibt sich durch Auswertung der Determinanten aus den zweiten gemischten Ableitungen nach den Variablen  $\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_{k+1} = \beta$ :

$$D_\nu = \begin{vmatrix} 2 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & 2 \end{vmatrix} = 2^\nu > 0$$

$(\nu \leq k)$

$$D_{k+1} = \begin{vmatrix} 2 & & & & & 2 n \mu_1 \\ & \ddots & & & & \vdots \\ & & \ddots & & & \vdots \\ & & & \ddots & & \vdots \\ 0 & & & & \ddots & 2 n \mu_k \\ & & & & & \vdots \\ 2 n \mu_1 & \dots & \dots & 2 n \mu_k & 2 \Sigma E(X_i^2) & \end{vmatrix} \quad \text{mit } \mu_i = E(\vartheta_i)$$

$$\begin{aligned}
 2^{-(k+1)} D_{k+1} &= \begin{vmatrix} 1 & & & n \mu_1 \\ & \ddots & & \vdots \\ & & 0 & \vdots \\ 0 & & & \vdots \\ & & & 1 & n \mu_k \\ n \mu_1 & \dots & n \mu_k & \Sigma E(X_i^2) \end{vmatrix} = \underbrace{\begin{vmatrix} 1 & & & n \mu_2 \\ & \ddots & & \vdots \\ & & \ddots & \vdots \\ & & & 1 & n \mu_k \\ n \mu_2 & \dots & n \mu_k & \Sigma E(X_i^2) \end{vmatrix}}_{= A_k} + \\
 &+ (-1)^k n \mu_1 \begin{vmatrix} 0 & \dots & 0 & n \mu_1 \\ 1 & & & \vdots \\ & \ddots & & \vdots \\ & & 0 & \vdots \\ 0 & & & \vdots \\ & & & 1 & n \mu_k \end{vmatrix} = A_k - (n \mu_1)^2 = A_{k-1} - \\
 &\quad - (n \mu_1)^2 - (n \mu_2)^2 = \dots \\
 &= A_2 - \sum_{i=1}^{k-1} (n \mu_i)^2 = \begin{vmatrix} 1 & n \mu_k \\ n \mu_k & \Sigma E(X_i^2) \end{vmatrix} - \sum_{i=1}^{k-1} (n \mu_i)^2 = \sum_1^k E(X_i^2) - \sum_1^k (n \mu_i)^2 = \\
 &= \sum_1^k E(X_i^2) - \sum_1^k E^2(X_i) = \sum_1^k \text{Var}(X_i) > 0
 \end{aligned}$$

**4.2 Verifizierung der Gleichung (13)**

Es gilt folgender Zusammenhang zwischen den Erwartungswerten und Varianzen bedingter Verteilungen (vgl. z. B. *M. H. De Groot 1970, S. 29*):

$$\text{Var}(X_i) = E \text{Var}(X_i | \vartheta) + \text{Var}[E(X_i | \vartheta)]$$

Daraus ergibt sich schrittweise

$$\begin{aligned}
 \text{Var}(X_i) &= E(n \vartheta_i - n \vartheta_i^2) + \text{Var}(n \vartheta_i) \\
 &= n E(\vartheta_i) - n E(\vartheta_i^2) + n^2 \text{Var}(\vartheta_i) \\
 &= n E(\vartheta_i) - n \text{Var}(\vartheta_i) - n E^2(\vartheta_i) + n^2 \text{Var}(\vartheta_i) \\
 &= n E(\vartheta_i) + n(n-1) \text{Var}(\vartheta_i) - n E^2(\vartheta_i) ,
 \end{aligned}$$

also mit  $\mu_i = E(\vartheta_i)$ ,  $\sigma_i^2 = \text{Var}(\vartheta_i)$

$$\Sigma \text{Var}(X_i) = n - n \Sigma \mu_i^2 + n(n-1) \Sigma \sigma_i^2$$

**4.3 Verifizierung von  $0 \leq \hat{\vartheta}_i \leq 1$**

Aus (14) folgt

$$(20) \quad \hat{\vartheta}_i = \frac{\mu_i (1 - \Sigma \mu_i^2 - \Sigma \sigma_i^2) + x_i \Sigma \sigma_i^2}{(1 - \Sigma \mu_i^2 - \Sigma \sigma_i^2) + n \Sigma \sigma_i^2}$$

Der Ausdruck

$$1 - \sum \mu_i^2 - \sum \sigma_i^2$$

ist wegen

$$\begin{aligned} \sum \sigma_i^2 &= \sum [E(\vartheta_i^2) - E^2(\vartheta_i)] \leq \sum [E(\vartheta_i) - E^2(\vartheta_i)] \\ &= 1 - \sum \mu_i^2 \end{aligned}$$

stets nichtnegativ. Infolgedessen ist auch  $\hat{\vartheta}_i$  nichtnegativ und kann den Wert 1 nicht überschreiten.

### Zusammenfassung

Bei der Vorgabe von  $k$  Einkommensgrößenklassen ist die gesamte Einkommensverteilung durch die  $k$  Anteilswerte festgelegt. In diesem Beitrag wird zunächst untersucht, welche Bayes'schen Schätzfunktionen sich bei Zugrundelegung verschiedener Schadensfunktionen und einer speziellen a priori Verteilung (nämlich einer *Dirichlet*-Verteilung) für die Anteilswerte ergeben. Danach wird eine (in den empirisch ermittelten Besetzungszahlen) lineare Bayes'sche Schätzfunktion angegeben. Diese Schätzfunktion setzt im wesentlichen nur die Kenntnis der a priori zu erwartenden Klassenanteile voraus und läßt sich als gewichtete Summe aus dem a priori zu erwartenden Klassenanteil und der Likelihood-Schätzfunktion darstellen.

### Summary

If the income scale is divided into  $k$  income ranges, the histogram form of the income distribution is completely determined by the  $k$  corresponding proportions. In section 2 Bayes estimators for these proportions are given for several loss functions and a special prior (*Dirichlet*-) distribution. The main purpose of this paper is the derivation of the linear Bayes estimator of section 3 without specifying a complete prior distribution. This linear Bayes estimator only depends on the first two moments of the prior distribution. It is shown that this estimator can be represented by a weighted average of the prior expected proportion and the maximum likelihood estimator.

### Literatur

- Bamberg, G. (1972), *Statistische Entscheidungstheorie*, Würzburg-Wien 1972.
- De Groot, M. H. (1970), *Optimal statistical decisions*, New York-St. Louis- San Francisco-London-Sydney-Toronto-Mexico-Panama 1970.
- Dickey, J. M. (1968), Smoothed estimates for multinomial cell probabilities, *Annals of Math. Statistics* 39 (1968), S. 561 - 566.
- Euler, M. (1974), Die Einkommensbefragung der privaten Haushalte, in: G. Fürst (Hrsg.): *Stand der Einkommensstatistik*, Göttingen 1974.
- Goldstein, M. (1975), A note on some Bayesian nonparametric estimates, *Annals of Statistics* 3 (1975), S. 736 - 740.
- Good, I. J. (1965), *The estimation of probabilities: An essay on modern Bayesian methods*, Cambridge (Mass.) 1965.

- Griffin, B. S. and R. G. Krutchkoff (1971)*, Optimal linear estimators: An empirical Bayes version with application to the binomial distribution, *Biometrika* 58 (1971), S. 195 - 201.
- Helten, E. (1971)*, Zur Bayes-Analyse, *Zeitschrift für Nationalökonomie und Statistik* 185 (1971), S. 528 - 545.
- Leonard, T. (1973)*, A Bayesian method for histograms, *Biometrika* 60 (1973), S. 297 - 308.
- Lidstone, G. J. (1920)*, Note to the general case of the Bayes-Laplace formula for inductive or a posteriori probabilities, *Trans. Fac. Actuar.* 8. (1920), S. 182 - 192.
- Pearson, K. (1934)*, Tables of the incomplete B-function, London 1934.
- Raiffa, H. and R. Schlaifer (1961)*, Applied statistical decision theory, Boston 1961.
- Salem, A. B. Z. and T. D. Mount (1974)*, A convenient descriptive model of income distribution: The Gamma density, *Econometrica* 42 (1974), S. 1115 bis 1127.
- Whittle, P. (1957)*, Curve and periodogram smoothing, *Journal of the Royal Statistical Society* 19 (1957), S. 38-47.