Gesetz der Großen Zahlen vs. Maximum Likelihood Estimation: Konvergenzgeschwindigkeit für Erwartungswerte

Matthias Isele (WV/AKS)

19.07.2023

württembergische

Ihr Fels in der Brandung.

Inhaltsverzeichnis

1 Gesetz der großen Zahlen

Theorem 1 (Gesetz der Großen Zahlen)**.** *Sei* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum und* $X := (X_i)_{i\in\mathbb{N}}$ *eine Folge von* $\mathcal{A}$*-messbaren Zufallsvariablen* $X_i : \Omega \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ *mit* $E(|X_i|) < \infty$ *für alle* $i \in \mathbb{N}$ *und*

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} (X_i - E(X_i)). \tag{1}$$

Die Folge X *erfüllt das* ***Schwache Gesetz der Großen Zahlen****, wenn für alle* $\varepsilon > 0$ *gilt*

$$\lim_{n\to\infty} P\big(\left|\overline{X}_n\right| > \varepsilon\big) = 0. \tag{2}$$

Die Folge X *erfüllt das* ***Starke Gesetz der Großen Zahlen****, wenn gilt*

$$P\big(\limsup_{n\to\infty} \left|\overline{X}_n\right| = 0\big) = 1. \tag{3}$$

Das Starke Gesetz d. großen Zahlen impliziert das Schwache Gesetz d. Großen Zahlen. Hinreichende Bedingungen an X *zur Erfüllung des Starken Gesetzes sind zum Beispiel:*

- *Die* X_i *sind paarweise unabhängig und identisch verteilt.*
- *Die* X_i *sind paarweise unkorreliert und die Folge ihrer Varianzen ist beschränkt.*

Erfüllt also die Folge von Zufallsvariablen $X := (X_i)_{i\in\mathbb{N}}$ das Starke Gesetz der großen Zahlen und haben die X_i identischen endlichen Erwartungswert $E(X_1) = E(X_2) = \ldots =: E(X)$, so konvergiert das Arithmetische Mittel der Folge punktweise fast überall gegen den Erwartungswert $E(X)$.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = E(X), \quad \textit{fast überall.} \tag{4}$$

Betrachte zum Beispiel die Familie $(\Omega_i, \mathcal{A}_i, P_i)_{i\in\mathbb{N}}$ aus identischen Kopien eines Wahrscheinlichkeitsraumes $(\Omega_0, \mathcal{A}_0, P_0)$ und eine Zufallsvariable X auf Ω_0 mit $E(|X|) < \infty$. Das Produktmodell

$$(\Omega, \mathcal{A}, P) := (\prod_{i\in\mathbb{N}} \Omega_i, \bigotimes_{i\in\mathbb{N}} \mathcal{A}_i, \bigotimes_{i\in\mathbb{N}} P_i) \tag{5}$$

erbt die Familie $(X_i)_{i\in\mathbb{N}}$ von Zufallsvariablen auf Ω definiert durch

$$X_i(\omega_1, \omega_2, \ldots) := X(\omega_i). \tag{6}$$

Die X_i sind per Konstruktion paarweise unabhängig und identisch verteilt. Es folgt nach Theorem 1, dass Gleichung (4) erfüllt ist. Interpretiert man die Zufallsvariable X als Beschreibung eines Zufallsexperimentes, so lässt sich deren Erwartungswert also beliebig gut durch hinreichend viele unabhängige Wiederholung des gleichen Zufallsexperimentes durch das Arithmetische Mittel approximieren. (Bis auf Fälle mit Wahrscheinlichkeit

Null.)

Im versicherungsmathematischen Umfeld möchte man Schadendurchschnitte schätzen, i.e. den Erwartungswert von Schadenhöhen bestimmen. Da in der Praxis niemals unendlich viele Stichproben zur Verfügung stehen, ist eine dringende Frage, wie schnell die Approxiamtion durch das arithmetische Mittel konvergiert und, ob es im Vergleich effektivere Schätzer gibt. Ein typisches Modell für Schadenhöhen ist die Lognormalverteilung. Im nächsten Abschnitt leiten wir eine Approximation für den Erwartungswert von lognormalverteilungen aus der Maximum Likelihood Methode her.

2 Maximum Likelihood Estimation für Lognormalverteilungen

Dieser Abschnitt folgt [1]. Sei $(\Omega_0, \mathcal{A}_0, P_0)$ ein Wahrscheinlichkeitsraum. Eine Zufallsvariable Y auf Ω_0 heißt lognormalverteilt, falls $Y = e^X$ mit einer (μ, σ)-normalverteilten Zufallsvariable X auf Ω_0. Das heißt, X hat die Wahrscheinlichkeitsverteilung

$$P(X \leq x) = \int_{-\infty}^{x} \mathcal{N}(\mu, \sigma)(z)\, dz \tag{7}$$

mit Dichte

$$\mathcal{N}(\mu, \sigma)(z) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{8}$$

Für die lognormalverteilte Zufallsvariable Y leite man die Wahrscheinlichkeitsverteilung

$$P(Y \leq y) = \int_{-\infty}^{y} g(\mu, \sigma)(z)\, dz \tag{9}$$

her mit Dichte

$$g(\mu, \sigma)(z) := \Theta(z) \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{z} e^{-\frac{(\ln(z)-\mu)^2}{2\sigma^2}}, \tag{10}$$

wobei $\Theta(z) = 1$ für $z > 0$ und $\Theta(z) = 0$ für $z \leq 0$ die Heaviside Stufenfunktion ist. Der Erwartungswert der Normalverteilung X ist $E(X) = \mu$. Naiverweise könnte man nun denken, dass der erwartungswert der Lognormalverteilung Y dann genau $E(Y) = e^{E(X)} = e^{\mu}$ ist. Dies ist aber falsch. Eine explizite Rechnung ergibt

$$E(Y) = \int_{-\infty}^{\infty} z\, g(\mu, \sigma)(z)\, dz = e^{\mu + \frac{\sigma^2}{2}}. \tag{11}$$

Das heißt, man braucht einen Korrekturfaktor $e^{\frac{\sigma^2}{2}}$. Wie im vorherigen Abschnitt betrachte man nun das Produktmodell

$$(\Omega, \mathcal{A}, P) := (\prod_{i \in \mathbb{N}} \Omega_i, \bigotimes_{i \in \mathbb{N}} \mathcal{A}_i, \bigotimes_{i \in \mathbb{N}} P_i) \tag{12}$$

aus Kopien von $(\Omega_0, \mathcal{A}_0, P_0)$ sowie die darauf definierte Familie $(Y_i)_{i\in\mathbb{N}}$ von lognormalverteilten Zufallsvariablen Y_i auf Ω mit

$$Y_i(\omega) = Y_i(\omega_1, \omega_2, \ldots) := Y(\omega_i). \tag{13}$$

Für festes $\omega \in \Omega$ ergibt sich eine Folge von unabhängigen Realisierungen

$$(y_i := Y_i(\omega))_{i\in\mathbb{N}}. \tag{14}$$

Betrachtet man nun die ersten n unabhängigen Realisierungen $\{y_1, \ldots, y_n\}$ der lognormalverteilten Zufallsvariable Y, so kann man Schätzer $(\hat{\mu}_n(\omega), \hat{\sigma}_n(\omega))$ für die Parameter (μ, σ) finden nach der Maximum Likelihood Methode:

$$(\hat{\mu}_n(\omega), \hat{\sigma}_n(\omega)) = \operatorname{argmax}_{\mu,\sigma} L(\mu, \sigma) \tag{15}$$

mit

$$L(\mu, \sigma) := \prod_{i=1}^{n} g(\mu, \sigma)(y_i). \tag{16}$$

Intuitiv werden also die Parameter (μ, σ) so geschätzt, dass die Realisierungen am wahrscheinlichsten sind. Man kann für Lognormalverteilungen zeigen, dass

$$\lim_{n\to\infty} (\hat{\mu}_n(\omega), \hat{\sigma}_n(\omega)) = (\mu, \sigma) \tag{17}$$

punktweise fast überall. Durch Nullsetzen der partiellen Ableitungen nch μ und σ in (16) ergibt sich

$$\hat{\mu}_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \ln(y_i), \quad \hat{\sigma}_n(\omega)^2 = \frac{1}{n}\sum_{i=1}^{n} (\ln(y_i) - \hat{\mu}_n(\omega))^2. \tag{18}$$

Einsetzen der Schätzer in den Erwartungswert (11) ergibt

$$\hat{E}_n(Y) := \sqrt[n]{\prod_{i=1}^{n} y_i}\, e^{\frac{1}{2n}\sum_{i=1}^{n}(\ln(y_i)-\mu_n)^2}, \quad \mu_n := \frac{1}{n}\sum_{i=1}^{n} \ln(y_i). \tag{19}$$

Der Schätzer setzt sich zusammen aus geometrischem Mittel

$$\operatorname{geom}(y_1, \ldots, y_n) := \sqrt[n]{\prod_{i=1}^{n} y_i} \tag{20}$$

und Korrekturfaktor

$$\operatorname{korr}(y_1, \ldots, y_n) := e^{\frac{1}{2n}\sum_{i=1}^{n}(\ln(y_i)-\mu_n)^2}. \tag{21}$$

Also,

$$\hat{E}_n(Y) = \operatorname{geom}(y_1, \ldots, y_n) \cdot \operatorname{korr}(y_1, \ldots, y_n). \tag{22}$$

Aufgrund der Konvergenz (17) gilt punktweise fast überall

$$\lim_{n\to\infty} \hat{E}_n(Y) = E(Y). \tag{23}$$

3 Inflation und Vergleich von Konvergenzgeschwindigkeiten

Aus den letzten beiden Abschnitten resultiert das folgende Theorem.

Theorem 2. *Sei Y eine lognormalverteilte Zufallsvariable mit $E(|Y|) < \infty$ und $y_1, \ldots, y_n$ unabhängige Realisierungen von Kopien von Y. Dann konvergieren sowohl das arithmetische Mittel*

$$arithm(y_1, \ldots, y_n) = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{24}$$

als auch das geometrische Mittel mit Korrekturfaktor

$$geom(y_1, \ldots, y_n) \cdot korr(y_1, \ldots, y_n) = \sqrt[n]{\prod_{i=1}^{n} y_i} \cdot e^{\frac{1}{2n} \sum_{i=1}^{n} (\ln(y_i) - \mu_n)^2}, \quad \mu_n := \frac{1}{n} \sum_{i=1}^{n} \ln(y_i), \tag{25}$$

für $n \to \infty$ gegen den Erwartungswert $E(Y)$, bis auf Fälle mit verschwindender Wahrscheinlichkeit.

Es ist verblüffend, dass solch scheinbar unterschiedliche Ausdrücke gegen den gleichen wert konvergieren. Dies liegt an der lognormalverteilung von Y. Für festes $a \geq 1$ betrachte man nun eine Inflation $Z = aY$. Da der Korrekturfaktor invariant unter Skalierungen ist,

$$\mathrm{korr}(y_1, \ldots, y_n) = \mathrm{korr}(ay_1, \ldots, ay_n), \tag{26}$$

ergeben sich für $a = \frac{E(Z)}{E(Y)}$ als mögliche Schätzer

$$\hat{a}_{\mathrm{geom},\, n} = \frac{\mathrm{geom}(z_1, \ldots, z_n)}{\mathrm{geom}(y_1, \ldots, y_n)} \tag{27}$$

und

$$\hat{a}_{\mathrm{arithm},\, n} = \frac{\mathrm{arithm}(z_1, \ldots, z_n)}{\mathrm{arithm}(y_1, \ldots, y_n)}. \tag{28}$$

Diese konvergieren beide gegen a aufgrund des obigen Theorems, bis auf Fälle mit verschwindender Wahrscheinlichkeit.

Da das arithmetische Mittel im Allgemeinen gegen den Erwartungswert konvergiert und das geometrische Mittel angepasst ist auf Lognormalverteilungen, ist zu erwarten, dass das geometrische Mittel schneller konvergiert für Lognormalverteilungen. Das heißt, es existiert ein $N \in \mathbb{N}$ sodass für alle $m \geq N$ gilt

$$\sum_{n=N}^{m} |\mathrm{geom}(y_1, \ldots, y_n) \cdot \mathrm{korr}(y_1, \ldots, y_n) - E(Y)|^2 < \sum_{n=N}^{m} |\mathrm{arithm}(y_1, \ldots, y_n) - E(Y)|^2. \tag{29}$$

Diese Vermutung konnte durch eine Simulation in Python bestätigt werden, siehe angehängtes Notebook.

Literatur

[1] M. Kohler, *Schadenversicherungsmathematik*. https://www.google.com/search?q=Schadenversicherungsmathematik+Prof.+Dr.+Michael+Kohler+Sommersemester+2009&rlz=1C1GCEB_enDE1041DE1041&oq=Schadenversicherungsmathematik+Prof.+Dr.+Michael+Kohler+Sommersemester+2009&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBBzU5OWowajeoAgCwAgA&sourceid=chrome&ie=UTF-8, Vorlesungsskript SoSe 2009, TU Darmstadt.