Normalapproximation für glatte Funktionen

Seminar Stochastische Geometrie und ihre Anwendungen

Tina Mayer | 11. April 2016 | Institut für Stochastik

Inhaltsverzeichnis

Warum beschäftigen wir uns mit Stein?

- Zentraler Grenzwertsatz für Summen unabhängiger Zufallsvariablen wird regelmäßig in der Statistik angewendet
- Klassischer Beweis beruht auf Fourier Methoden
- dieser kann keine Aussage über die Konvergenzgeschwindigkeit machen
- zudem liegen in der Anwendung meist abhängige Zufallsvariablen vor

Warum beschäftigen wir uns mit Stein?

- Problem: Fourier-Methode ist schwerer anzuwenden und Grenzen für die Genauigkeit der Approximation sind schwerer zu finden
- Lösung: Systematische Methodik, die für unabhängige Zufallsvariablen anwendbar ist und zudem auch Abschätzungen für die Approximationsgenauigkeit angibt.

Einführung Abstände

- Um analysieren zu können, wie nahe die Verteilung für festes n an der Normalverteilung ist, können verschiedene Abstände verwendet werden.
- Diese unterscheiden sich durch die Eigenschaften der Funktionen.
- Im Folgenden werden einige wichtige Abstände eingeführt.

Totalvariationsnorm

$$d_{TV}(P,Q) := \sup_{h \in H} \left| \int h \, dP - \int h \, dQ \right| = \sup_{A} \left| P(A) - Q(A) \right|$$

wobei H = $\{\mathbb{1}_A : A \textit{ messbar}\}$

Kolmogorov-Abstand

$$\begin{aligned} d_K(P, Q) &= \sup_{h \in H} \left| \int h \, dP - \int h \, dQ \right| \\ &= \sup_{z \in \mathbb{R}} |P(-\infty, z] - Q(-\infty, z]| \end{aligned}$$

wobei H = $\{\mathbb{1}_{(-\infty,z]} : z \in \mathbb{R}\}$

Wasserstein-Abstand

$$d_W(P, Q) := \sup_{h \in H} \left| \int h \, dP - \int h \, dQ \right|$$

wobei H = $\{h : \mathbb{R} \longrightarrow \mathbb{R};\ \| h' \| \leqslant 1\} =: Lip(1)$ und
g: $\mathbb{R} \longrightarrow \mathbb{R}$, $\|g\|$ bezeichnet $\sup_{x \in \mathbb{R}} |g(x)|$

beschränkter Wasserstein-Abstand

$$d_{BW(k)}(P,Q) := \sup_{h \in H} \left| \int h \, dP - \int h \, dQ \right|$$

wobei H = $\{h : \mathbb{R} \longrightarrow \mathbb{R};\ \| h \| \leqslant 1, \| h' \| \leqslant k\}$, das heißt H ist die Menge der durch 1 beschränkten Lipschitzstetigen Funktionen mit Lipschitzkonstante $\leq k$.

Charakterisierungen

Im Folgenden gilt:
$Z \sim N(0,1), \mathcal{C}_{b,d} := \{f : \mathbb{R} \longrightarrow \mathbb{R}$, f stetig und stückweise stetig differenzierbar und $\mathbb{E}|f'(Z)| < \infty$.

Die Stein-Methode wird durch folgendes Lemma motiviert:

Lemma 1

Sei W eine reellwertige Zufallsvariable. Dann ist $W \sim N(0,1)$ genau dann, wenn für alle $f \in \mathcal{C}_{b,d}$ mit $\mathbb{E}|f'(W)| < \infty$ folgendes gilt:

$$\mathbb{E}f'(W) = \mathbb{E}\{Wf(W)\} \tag{1}$$

Beweis Lemma 1

W∼N(0,1) ⇒ (1): nachrechnen

Sei W eine standardnormalverteilte Zufallsvariable. Dann gilt für $f \in \mathcal{C}_{b,d}$:

$$
\begin{aligned}
\mathbb{E}f'(W) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(w) e^{-w^2/2} dw \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(w) \left(\int_{-\infty}^{w} (-x) e^{-x^2/2} dx \right) dw \\
&+ \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} f'(w) \left(\int_{w}^{\infty} x e^{-x^2/2} dx \right) dw \\
&\overset{Fubini}{=} \ldots = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [f(x) - f(0)] x e^{-x^2/2} dx \\
&= \mathbb{E} W f(W)
\end{aligned}
$$

Beweis Lemma 1

(1) $\Rightarrow$ W$\sim$N(0,1):

Für festes $z \in \mathbb{R}$ bezeichnet $f(w) := f_z(w)$ die Lösung der folgenden Gleichung

$$f'(w) - wf(w) = \mathbb{1}_{(-\infty,z]}(w) - \Phi(z) \tag{2}$$

Multipliziert man (2) mit $-e^{-w^2/2}$ auf beiden Seiten erhält man

$$(e^{-w^2/2}f(w))' = -e^{-w^2/2}(\mathbb{1}_{(-\infty,z]}(w) - \Phi(z))$$

Beweis Lemma 1

Daher is f_z gegeben durch

$$f_z(w) = e^{w^2/2} \int_{-\infty}^{w} (\mathbb{1}_{(-\infty,z]}(x) - \Phi(z)) e^{-x^2/2} dx$$
$$= \begin{cases} \sqrt{2\pi} e^{w^2/2} \Phi(w)[1 - \Phi(z)] \text{ falls } w \leq z \\ \sqrt{2\pi} e^{w^2/2} \Phi(z)[1 - \Phi(w)] \text{ falls } w \geq z \end{cases} \tag{3}$$

Beweis Lemma 1

f_z ist eine beschränkte, stetige und stückweise stetig differenzierbare Funktion. Gelte (1) für alle $f \in \mathcal{C}_{b,d}$. Dann gilt (1) auch für f_z und daher gilt mit (2):

$$\begin{aligned} 0 &= \mathbb{E}[f_z'(W) - Wf_z(W)] \\ &= \mathbb{E}[1_{(-\infty,z]}(w) - \Phi(z)] \\ &= P(W \leq z) - \Phi(z) \end{aligned}$$

Daher ist W standardnormalverteilt. $\square$

Bemerkung

Gleichung (2) ist ein Spezialfall der Stein-Gleichung

$$f'(w) - wf(w) = h(w) - \mathbb{E}h(Z) \tag{4}$$

welche nach f zu lösen ist, wobei h eine reellwertige, messbare Funktion mit $\mathbb{E}|h(Z)| < \infty$ darstellt.
Die Lösung $f = f_h$ für (4) ist gegeben durch

$$f_h(w) = -e^{w^2/2} \int_w^\infty [h(x) - \mathbb{E}h(Z)] e^{-x^2/2} dx \tag{5}$$

wobei hier ähnlich wie bei (3) vorgegangen wird.

Eigenschaften der Lösungen für die Stein-Gleichungen

Im folgenden werden Eigenschaften der Lösung (5) der Stein-Gleichungen (4) aufgelistet. Wir benötigen diese um im Folgenden die Fehlergrenzen für die verschiedenen Approximationen abschätzen zu können.

Lemma 2

Für jede absolut stetige Funktion $h : \mathbb{R} \to \mathbb{R}$ erfüllt die Lösung f_h der Stein-Gleichung (4) folgende Eigenschaften:

$$\|f_h\| \leq \min\left(\sqrt{\frac{\pi}{2}}\|h(\cdot) - \mathbb{E}h(Z)\|, 2\|h'\|\right)$$
$$\|f_h'\| \leq \min(2\|h(\cdot) - \mathbb{E}h(Z)\|, 4\|h'\|)$$
$$\|f_h''\| \leq 2\|h'\| \tag{6}$$

[ohne Beweis]

Konstruktion der Stein-Gleichungen

Seien $\xi_1, \xi_2, ...\xi_n$ unabhängige Zufallsvariablen, die $\mathbb{E}\xi_i = 0$ für $1 \leq i \leq n$ und $\sum_{i=1}^{n} \mathbb{E}\xi_i^2 = 1$ erfüllen. Dabei müssen die ξ_i nicht identisch verteilt sein. Wir setzten

$$W = \sum_{i=1}^{n} \xi_i \text{ und } W^{(i)} = W - \xi_i$$

und definieren

$$K_i(t) = \mathbb{E}\{\xi_i(I_{\{0 \leq t \leq \xi_i\}} - I_{\{\xi_i \leq t < 0\}})\}$$

Zudem gilt $K_i(t) \geq 0$ für alle reellen Zahlen t, sowie

$$\int_{-\infty}^{\infty} K_i(t)dt = \mathbb{E}\xi_i^2 \text{ und } \int_{-\infty}^{\infty} |t|K_i(t)dt = \frac{1}{2}\mathbb{E}|\xi_i^3|$$

Konstruktion der Stein-Gleichungen

Sei h eine messbare Funktion mit $\mathbb{E}|h(Z)| < \infty$ und $f = f_h$ sei die Lösung der Stein-Gleichung (4). Das Ziel ist es

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}\{f'(W) - Wf(W)\}$$

abzuschätzen.

Für das weitere Vorgehen ist folgender Ansatz fundamental:

Konstruktion der Stein-Gleichungen

$$
\begin{aligned}
\mathbb{E}\{Wf(W)\} &= \sum_{i=1}^{n} \mathbb{E}\{\xi_i f(W)\} \\
&= \sum_{i=1}^{n} \mathbb{E}\{\xi_i [f(W) - f(W^{(i)})]\} \\
&= \sum_{i=1}^{n} \mathbb{E}\{\xi_i \int_0^{\xi_i} f'(W^{(i)} + t) dt\} \\
&= \sum_{i=1}^{n} \mathbb{E}\{\int_{-\infty}^{\infty} f'(W^{(i)} + t)\xi_i (I_{\{0 \le t \le \xi_i\}} - I_{\{\xi_i \le t < 0\}}) dt\} \\
&= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \mathbb{E}\{f'(W^{(i)} + t)\} K_i(t) dt \qquad (7)
\end{aligned}
$$

Konstruktion der Stein-Gleichungen

Aus der Definition von K_i und $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ folgt

$$\mathbb{E}f'(W) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \mathbb{E}\{f'(W)\}K_i(t)dt \tag{8}$$

Daher folgt aus (7) und (8)

$$\mathbb{E}\{f'(W) - Wf(W)\} = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \mathbb{E}\{f'(W) - f'(W^{(i)} + t)\}K_i(t)dt$$

Ziel

Das Ziel ist es $\mathbb{E}h(W) - \mathbb{E}h(Z)$ für verschiedene Klassen von Zufallsvariablen W abzuschätzen, wobei h eine ausreichend glatte Funktion ist, welche

$$\|h'\| := \sup_x |h'(x)| < \infty$$

erfüllt.

Theorem 1

Wir nehmen an, dass ein δ existiert, so dass für jede Lipschitzstetige Funktion h

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \delta \|h'\|$$

gilt. Dann folgt

$$d_W(L(W), N(0,1)) := \sup_{h \in Lip(1)} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \delta$$

$$d_K(L(W), N(0,1)) := \sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 2\delta^{1/2}$$

[ohne Beweis]

Ziel

Im Folgenden wird gezeigt, dass Theorem 1 mit passendem δ auch erfüllt ist, wenn W eine Summe unabhängig verteilter Zufallsvariablen oder lokal abhängiger Zufallsvariablen ist. Dabei müssen die Zufallsvariablen nicht identisch verteilt sein!

Theorem 2

Seien $\xi_1, \xi_2, ..., \xi_n$ unabhängige Zufallsvariablen, die $\mathbb{E}\xi_i = 0$ und $\mathbb{E}|\xi_i|^3 < \infty$ für $1 < i \leq n$ erfüllen, sodass $\sum_{i=1}^{n} \mathbb{E}\xi_i^2 = 1$ gilt. Dann kann Theorem 1 angewendet werden mit

$$\delta = 3 \sum_{i=1}^{n} \mathbb{E}|\xi_i|^3 \tag{9}$$

Es gilt also

$$|\mathbb{E}|W| - \sqrt{\frac{2}{\pi}}| \leq 3 \sum_{i=1}^{n} \mathbb{E}|\xi_i|^3$$

Beweis Theorem 2

Mit (6), (8), (9) und dem Mittelwertsatz folgt:

$$
\begin{aligned}
&|\mathbb{E}\{f_h'(W) - Wf_h(W)\}| \\
&\qquad \leq \sum_{i=1}^{\infty} \mathbb{E}|f_h'(W) - f_h'(W^{(i)} + t)|K_i(t)dt \\
&\qquad \leq 2\|h'\| \sum_{i=1}^{n} \int_{-\infty}^{\infty} \mathbb{E}(|t| + |\xi_i|)K_i(t)dt \\
&\qquad \leq 2\|h'\| \sum_{i=1}^{n} (\frac{\mathbb{E}|\xi_i|^3}{2} + \mathbb{E}|\xi_i|\mathbb{E}\xi_i^2) \\
&\qquad \leq 3\|h'\| \sum_{i=1}^{n} \mathbb{E}|\xi_i|^3
\end{aligned}
$$

Theorem 3

Seien $\xi_1, \xi_2, ..., \xi_n$ unabhängige Zufallsvariablen, die $\mathbb{E}\xi_i = 0$ für alle $1 \leq i \leq n$ erfüllen, sodass $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ gilt. Dann kann Theorem 1 angewendet werden mit

$$\delta = 4(4\beta_2 + 3\beta_3)$$

wobei

$$\beta_2 = \sum_{i=1}^n \mathbb{E}\xi_i^2 I_{\{|\xi_i|>1\}} \textit{ und } \beta_3 = \sum_{i=1}^n \mathbb{E}|\xi_i|^3 I_{\{|\xi_i|\leq 1\}}$$

[ohne Beweis]

Korollar: Zentraler Grenzwertsatz von Lindeberg

Theorem 1 und Theorem 3 zusammen liefern den Zentralen Grenzwertsatz von Lindeberg wie folgt:
$\forall n \in \mathbb{N}$ sei $X_{n1}, ..., X_{nn} : \Omega \longrightarrow \mathbb{R}$ unabhängige Zufallsvariablen mit

$$\mathbb{E}X_{nk} = 0$$

$$0 < \sigma_{nk}^2 = Var(X_{nk}) < \infty$$

für $k \in \{1, ..., n\}$ und $\sum_{k=1}^{n} \sigma_{nk}^2 = 1$. Für jedes $\varepsilon > 0$ gelte

$$\lim_{n \longrightarrow \infty} \sum_{k=1}^{n} \mathbb{E}|X_{nk}|^2 \mathbb{1}_{\{|X_{nk}| > \varepsilon\}} = 0. \tag{10}$$

Dann gilt:

$$\lim_{n \to \infty} \mathbb{P}(X_{n1} + ... + X_{nn} \leq x) = \Phi(x) \; \forall x \in \mathbb{R}.$$

Beweis Satz von Lindeberg

Verwende Theorem 3 mit $\xi_i := X_{ni},\ i = 1, 2, ..., n$. Nach Voraussetzung gilt:

$$\mathbb{E}\xi_i = \mathbb{E}X_{ni} = 0 \quad \text{und}$$
$$\sum_{i=1}^{n} \mathbb{E}\xi_i^2 = \sum_{i=1}^{n} \sigma_{ni}^2 = 1,\ i = 1, ..., 2.$$

Daher ist Theorem 3 anwendbar und es ist

$$\delta = 4(4\beta_2 + 3\beta_3) \leq 16(\beta_2 + \beta_3)$$
$$= \sum_{i=1}^{n} \mathbb{E}X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>1\}} + \sum_{i=1}^{n} \mathbb{E}|X_{ni}|^3 \mathbb{1}_{\{|X_{ni}|\leq 1\}}$$

Der erste Summand konvergiert wegen (10) für $n \longrightarrow \infty$ gegen 0.

Beweis Satz von Lindeberg

Für den zweiten Summand und $\varepsilon \in (0,1)$ beliebig gilt:

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}|X_{ni}|^3 \mathbb{1}_{\{|X_{ni}|\leq 1\}} \\
&\qquad = \sum_{i=1}^{n} \mathbb{E}|X_{ni}^3 \mathbb{1}_{\{|X_{ni}|\leq \varepsilon\}} + \sum_{i=1}^{n} \mathbb{E}|X_{ni}|^3 \mathbb{1}_{\{\varepsilon < |X_{ni}|\leq 1\}} \\
&\qquad \leq \varepsilon \sum_{i=1}^{n} \mathbb{E}|X_{ni}|^2 + \sum_{i=1}^{n} \mathbb{E}|X_{ni}|^2 \mathbb{1}_{\{|X_{ni}| > \varepsilon\}} \\
&\qquad = \varepsilon \cdot 1 + \sum_{i=1}^{n} \mathbb{E}|X_{ni}|^2 \mathbb{1}_{\{|X_{ni}| > \varepsilon\}},
\end{aligned}
$$

wobei auch hier der zweite Summand wegen (10) für $n \longrightarrow \infty$ gegen 0 konvergiert.

Beweis Satz von Lindeberg

$\Longrightarrow \delta \leq \varepsilon$. Da ε beliebig klein gewählt werden kann, gilt also $\lim_{n \longrightarrow \infty} \delta = 0$ und somit gilt wegen Theorem 1 die Behauptung. $\square$

Definition m-abhängig

Eine m-abhängige Folge von Zufallsvariablen $\xi_i, i \in \mathbb{Z}$ hat die Eigenschaft, dass für jedes i die Mengen der Zufallsvariablen $\{\xi_j, j \leq i\}$ und $\{\xi_i, j > i + m\}$ unabhängig sind.

Ein Spezialfall sind 0-abhängige Folgen von Zufallsvariablen.

LD

Sei J eine endliche Indexmenge der Kardinalität n und sei $\{\xi_i, i \in J\}$ ein Indexfeld mit Mittelwert 0 und endlicher Varianz. Definiere $W = \sum_{i \in J} \xi_i$ und nehme an, dass $Var(W) = 1$ gilt. Für $A \subset J$ bezeichne $\{\xi_i, i \in A\}$ und $A^c = \{j \in J : j \notin A\}$. Dann gilt

(LD) Für alle $i \in J$ existiert ein $A_i \subset J$ sodass ξ_i und $\xi_{A_i^c}$ unabhängig sind.

Theorem 4

Theorem 1 kann mit

$$\delta = 4\mathbb{E}|\sum_{i\in J}\{\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i)\}| + \sum_{i\in J}\mathbb{E}|\xi_i\eta_i^2|$$

unter (LD) angewendet werden.

Bemerkung:
Für unabhängige Zufallsvariablen ist der Wert von δ in $5\sum_{i\in J}\mathbb{E}|\xi_i|^3$ etwas größer als die Beschränkung, die in Theorem 2 gegeben ist.

Literatur

A.D. Barbour, Louis H. Y. Chen. *An Introduction to Stein's Method.* Singapore University Press, Singapore, 2005 (pp.1-19).

I. Nourdin, G. Peccati. *Normal Approximations with Malliavin Calculus - From Stein's Method to Universality.* Cambridge University Press, New Yourk, 2012.

Prof. Dr. V. Schmidt. *Vorlesungsskript Elementare Wahrscheinlichkeitsrechnung und Statistik.* Institut für Stochastik. Ulm, 2011.