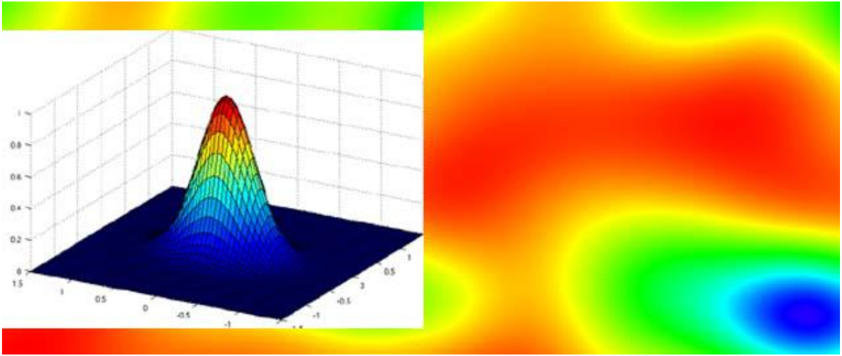

Die Methode von Stein für multivariate Normalverteilung

Seminar zur Methode von Stein

Stefan Roth | Institut für Stochastik | 6. Juni, 2016

Universität Ulm

Inhalt

Motivation

- Lemma von Stein: Sei Z eine reellwertige Zufallsvariable. Dann gilt:

$$Z \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad \mathbb{E}[f'(Z) - Zf(Z)] = 0,$$

für jede stetige und stückweise stetig differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ ($f \in \mathcal{C}_{bd}$) mit $\mathbb{E}\,|f'(Z)| < \infty$.

- Stein Gleichung: Seien $Z \sim \mathcal{N}(0,1)$ und $h : \mathbb{R} \to \mathbb{R}$ messbar, so, dass $\mathbb{E}\,|h(Z)| < \infty$. Dann existiert eine Funktion $f_h \in \mathcal{C}_{bd}$ mit

$$f_h'(w) - wf_h(w) = h(w) - \mathbb{E}\,h(Z), \quad \forall\, w \in \mathbb{R}.$$

$\rightsquigarrow$ Ziel: Verallgemeinerung für Zufallsvektoren.

Notation

- $\langle \cdot, \cdot \rangle_{\mathbb{R}^d}$, $\| \cdot \|_{\mathbb{R}^d}$ - euklidisches Skalarprodukt und euklidische Norm in $\mathbb{R}^d$
- $\mathcal{M}_d(\mathbb{R})$ - reellwertige $d \times d$ - Matrizen
- $A, B \in \mathcal{M}_d(\mathbb{R})$:

  $$\langle A, B \rangle_{HS} = \mathrm{Tr}(AB^\top), \quad \|A\|_{HS} = \sqrt{\langle A, A \rangle_{HS}},$$

  Hilbert-Schmidt Skalarprodukt und Hilbert-Schmidt Norm.
- $A \in \mathcal{M}_d(\mathbb{R})$:

  $$\|A\|_{op} = \sup\{\|Ax\|_{\mathbb{R}^d} : \; x \in \mathbb{R}^d, \; \|x\|_{\mathbb{R}^d} = 1\},$$

  Operatornorm von A.

Multivariates Lemma von Stein

Seien

- $C \in \mathcal{M}_d(\mathbb{R})$ positiv semi-definit und symmetrisch
- $N = (N_1, \ldots, N_d)$ ein $\mathbb{R}^d$-wertiger Zufallsvektor

Dann gilt: N ist genau dann ein Gausscher Zufallsvektor mit Kovariantmatrix C, wenn

$$\mathbb{E}\left[\langle N, \nabla f(N)\rangle_{\mathbb{R}^d}\right] = \mathbb{E}\left[\langle C, \mathrm{H}_f(N)\rangle_{HS}\right] \tag{1}$$

für jede $\mathcal{C}^2$-Funktion $f : \mathbb{R}^d \to \mathbb{R}$ mit beschränkter erster und zweiter Ableitung.

Bemerkung:

- $\mathcal{C}^2 = \{f : \mathbb{R}^d \to \mathbb{R} : \ f \text{ zweimal stetig diff'bar}\}$
- $\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d}\right)^\top$, $\mathrm{H}_f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{1 \le i,j \le d}$

Beweisidee: Notwendigkeit von (1): Sei $N \sim \mathcal{N}(0, C)$, $C = (c_{ij})_{1 \leq i,j \leq d}$. Dann gilt

$$\mathbb{E}[N_i f(N_1, \ldots, N_d)] = \sum_{j=1}^{d} c_{ij}\, \mathbb{E}\left[\frac{\partial f}{\partial x_j}(N_1, \ldots, N_d)\right],$$

für jedes $i \in \{1, \ldots, d\}$, und damit

$$\begin{aligned}
\mathbb{E}\left[\langle N, \nabla f(N)\rangle_{\mathbb{R}^d}\right] &= \sum_{i=1}^{d} \mathbb{E}\left[N_i \frac{\partial f}{\partial x_i}(N_1, \ldots, N_d)\right] \\
&= \sum_{i=1}^{d} \sum_{j=1}^{d} c_{ij}\, \mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}(N_1, \ldots, N_d)\right] \\
&= \mathbb{E}\left[\langle C, \mathrm{H}_f(N)\rangle_{HS}\right]
\end{aligned}$$

Hinlänglichkeit von (1): Sei $G \sim \mathcal{N}(0, C)$ unabhängig von N sowie $f : \mathbb{R}^d \to \mathbb{R}$ wie im Lemma. Es genügt zu zeigen:

$$\mathbb{E}[f(N)] - \mathbb{E}[f(G)] = 0,$$

für alle $f \in \mathcal{C}^2$ mit beschränkten Ableitungen. Für $\varphi(t) = \mathbb{E}\left[f\left(\sqrt{t}N + \sqrt{1-t}G\right)\right]$ gilt wegen

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt,$$

dass

$$\begin{aligned}
&\mathbb{E}[f(N)] - \mathbb{E}[f(G)] \\
&= \int_0^1 \mathbb{E}\left[\mathbb{E}\left[\left\langle \nabla f\left(\sqrt{t}N + \sqrt{1-t}x\right), N\right\rangle_{\mathbb{R}^d}\right]\Big|_{x=G}\right] \frac{dt}{2\sqrt{t}} \\
&- \int_0^1 \mathbb{E}\left[\mathbb{E}\left[\left\langle \nabla f\left(\sqrt{t}x + \sqrt{1-t}G\right), G\right\rangle_{\mathbb{R}^d}\right]\Big|_{x=N}\right] \frac{dt}{2\sqrt{1-t}}
\end{aligned}$$

Wegen (1) gilt:

$$\begin{aligned} h_1(x) &:= \mathbb{E}\left[\left\langle \nabla f\left(\sqrt{t}N + \sqrt{1-t}x\right), N\right\rangle_{\mathbb{R}^d}\right] \\ &= \sqrt{t}\,\mathbb{E}\left[\left\langle C, \mathrm{H}_f(\sqrt{t}N + \sqrt{1-t}x)\right\rangle_{HS}\right] \end{aligned}$$

sowie

$$\begin{aligned} h_2(x) &:= \mathbb{E}\left[\left\langle \nabla f\left(\sqrt{t}x + \sqrt{1-t}G\right), G\right\rangle_{\mathbb{R}^d}\right] \\ &= \sqrt{1-t}\,\mathbb{E}\left[\left\langle C, \mathrm{H}_f(\sqrt{t}x + \sqrt{1-t}G)\right\rangle_{HS}\right] \end{aligned}$$

Damit folgt

$$\mathbb{E}[f(N)] - \mathbb{E}[f(G)] = 0,$$

für alle $f \in \mathcal{C}^2$ mit beschränkten Ableitungen. $\square$

Stein Gleichung für $\mathcal{N}(0, I_d)$

Sei $N \sim \mathcal{N}(0, I_d)$ und $h : \mathbb{R}^d \to \mathbb{R}$ mit $\mathbb{E}\,|h(N)| < \infty$.

Stein Gleichung:

$$\Delta f(x) - \langle x, \nabla f(x) \rangle_{\mathbb{R}^d} = h(x) - \mathbb{E}[h(N)], \quad x \in \mathbb{R}^d.$$

Eine Funktion f heißt Lösung der Stein Gleichung, falls sie in $\mathcal{C}^2$ liegt und vorherige Gleichung für jedes $x \in \mathbb{R}^d$ löst.

Theorem 1: Ist $h : \mathbb{R}^d \to \mathbb{R}$ eine Lipschitzfunktion mit L-Konstante $K = \sup_{x \neq y} \frac{|h(x)-h(y)|}{||x-y||_{\mathbb{R}^d}} > 0$, so ist

$$f_h(x) = \int_0^\infty \mathbb{E}\left[h(N) - h(e^{-t}x + \sqrt{1-e^{-2t}}N)\right] dt,$$

$x \in \mathbb{R}^d$, eine Lösung der Stein Gleichung. Außerdem gilt:

$$\sup_{x \in \mathbb{R}^d} ||\mathrm{H}_{f_h}(x)||_{HS} \leq \sqrt{d}K.$$

Stein Gleichung für positiv definite Kovarianzmatrizen

Sei nun $C \in \mathcal{M}_d(\mathbb{R})$ eine positiv definite Kovarianzmatrix und $N \sim \mathcal{N}(0, C)$ sowie $h : \mathbb{R}^d \to \mathbb{R}$ so, dass $\mathbb{E}\,|h(N)| < \infty$.

Stein Gleichung:

$$\langle C, \mathrm{H}_f(x) \rangle_{HS} - \langle x, \nabla f(x) \rangle_{\mathbb{R}^d} = h(x) - \mathbb{E}[h(N)], \quad x \in \mathbb{R}^d.$$

Eine Funktion f heißt Lösung der Stein Gleichung, falls sie in $\mathcal{C}^2$ liegt und vorherige Gleichung für jedes $x \in \mathbb{R}^d$ löst.

Theorem 2: Ist $h : \mathbb{R}^d \to \mathbb{R}$ eine Lipschitzfunktion mit L-Konstante $K > 0$, so ist

$$f_h(x) = \int_0^\infty \mathbb{E}\left[h(N) - h(e^{-t}x + \sqrt{1 - e^{-2t}}N)\right] dt,$$

$x \in \mathbb{R}^d$, eine Lösung der Stein Gleichung. Außerdem gilt:

$$\sup_{x \in \mathbb{R}^d} ||\mathrm{H}_{f_h}(x)||_{HS} \leq \sqrt{d}K||C^{-1}||_{op}||C||_{op}^{1/2}.$$

Bemerkung: Ist $C \in \mathcal{M}_d(\mathbb{R})$ positiv definit, so gibt es eine positiv definite Matrix $A \in \mathcal{M}_d(\mathbb{R})$ mit $C = A^2$.

Beweisidee: Sei $A \in \mathcal{M}_d(\mathbb{R})$ wie zuvor und

$$V(x) = \int_0^\infty \mathbb{E}[h_A(A^{-1}N) - h_A(e^{-t}x + \sqrt{1-e^{-2t}}A^{-1}N)]dt,$$

mit $h_A(x) = h(Ax)$. Dann gilt $f_h(x) = V(A^{-1}x)$. Wegen $A^{-1}N \sim \mathcal{N}(0, I_d)$ folgt aus Theorem 1

$$\Delta V(x) - \langle x, \nabla V(x) \rangle_{\mathbb{R}^d} = h_A(x) - \mathbb{E}[h_A(Y)],$$

für alle $x \in \mathbb{R}^d$ und für $Y \sim \mathcal{N}(0, I_d)$. Es gilt:

$$\begin{aligned} \Delta V(x) &= \left\langle C, \mathrm{H}_{f_h}(Ax) \right\rangle_{HS} \\ \langle x, \nabla V(x) \rangle_{\mathbb{R}^d} &= \langle Ax, \nabla f_h(Ax) \rangle_{\mathbb{R}^d} \\ \mathbb{E}[h_A(Y)] &= \mathbb{E}[h(N)] \end{aligned}$$

Wegen $h_A(x) = h(Ax)$ also

$$\begin{aligned}\left\langle C, \mathrm{H}_{f_h}(Ax)\right\rangle_{HS} - \langle Ax, \nabla f_h(Ax)\rangle_{\mathbb{R}^d} &= \Delta V(x) - \langle x, \nabla V(x)\rangle_{\mathbb{R}^d} \\ &= h_A(x) - \mathbb{E}[h_A(Y)] \\ &= h(Ax) - \mathbb{E}[h(N)]\end{aligned}$$

Mit der Substitution $x \mapsto A^{-1}x$ folgt

$$\left\langle C, \mathrm{H}_{f_h}(x)\right\rangle_{HS} - \langle x, \nabla f_h(x)\rangle_{\mathbb{R}^d} = h(x) - \mathbb{E}[h(N)]$$

Bleibt die Abschätzung

$$\sup_{x \in \mathbb{R}^d} ||\mathrm{H}_{f_h}(x)||_{HS} \leq \sqrt{d} K ||C^{-1}||_{op} ||C||_{op}^{1/2}$$

zu zeigen.

Zunächst folgt aus der Abschätzung von Theorem 1:

$$\sup_{x \neq y} ||\mathrm{H}_V(x)||_{HS} \leq \sqrt{d} K ||A||_{op}$$

Außerdem:

$$\begin{aligned}
\sup_{x \in \mathbb{R}^d} ||\mathrm{H}_{f_h}(x)||_{HS} &= \sup_{x \in \mathbb{R}^d} ||A^{-1} \mathrm{H}_V(A^{-1}x) A^{-1}||_{HS} \\
&= \sup_{x \in \mathbb{R}^d} ||A^{-1} \mathrm{H}_V(x) A^{-1}||_{HS} \\
&\leq ||A^{-2}||_{op} \sup_{x \in \mathbb{R}^d} ||\mathrm{H}_V(x)||_{HS} \\
&\leq ||A^{-2}||_{op} \sqrt{d} K ||A||_{op} \\
&= ||C^{-1}||_{op} ||C||_{op}^{1/2} \sqrt{d} K
\end{aligned}$$

$\square$

Schranken für den Wassersteinabstand

Sei $\mathcal{H}$ die Menge der Lipschitzfunktionen $h : \mathbb{R}^d \to \mathbb{R}$ mit Lipschitzkonstante 1 und seien F, G Zufallsvektoren mit $\mathbb{E}[h(F)], \mathbb{E}[h(G)] < \infty, \forall\, h \in \mathcal{H}$.

Wassersteinabstand:

$$d_W(F, G) = \sup_{h \in \mathcal{H}} |\, \mathbb{E}[h(F)] - \mathbb{E}[h(G)]|$$

Theorem 3: Sei $C \in \mathcal{M}_d(\mathbb{R})$ positiv definit und $N \sim \mathcal{N}(0, C)$. Dann gilt für jeden quadratisch integrierbaren Zufallsvektor F

$$d_W(F, N) \leq \sup_{f \in \mathcal{F}_W^d(C)} |\mathbb{E}\langle C, \mathrm{H}_f(F)\rangle_{HS} - \mathbb{E}\langle F, \nabla f(F)\rangle_{\mathbb{R}^d}|,$$

wobei $\mathcal{F}_W^d(C) = \mathcal{C}^2 \cap \{f : \mathbb{R}^d \to \mathbb{R} : \sup_{x \in \mathbb{R}^d} ||\mathrm{H}_f(x)||_{HS} \leq \sqrt{d}||C^{-1}||_{op}||C||_{op}^{1/2}\}$

Beweis: Es gilt

$$\begin{aligned} d_W(F,N) &= \sup_{h\in\mathcal{H}} |\mathbb{E}[h(F)] - \mathbb{E}[h(N)]| \\ &= \sup_{h\in Lip(1)} |\mathbb{E}\langle C, \mathrm{H}_{f_h}(F)\rangle_{HS} - \mathbb{E}\langle F, \nabla f_h(F)\rangle_{\mathbb{R}^d}| \\ &\leq \sup_{f\in\mathcal{F}_W^d(C)} |\mathbb{E}\langle C, \mathrm{H}_f(F)\rangle_{HS} - \mathbb{E}\langle F, \nabla f(F)\rangle_{\mathbb{R}^d}|, \end{aligned}$$

wobei die zweite Gleichheit wegen Theorem 2 gilt.

$\square$

Literatur

[1] A.D. Barbour, Louis H. Y. Chen. *An Introduction to Stein's Method.* Singapore University Press, Singapore, 2005 (pp. 23-39).

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