Ein schneller Schätzer zur Berechnung der Kovarianzfunktion: Spektralansatz

S. Wilke
November 2002

1 Notation

- Ein **Zufallsfeld** ist eine Familie von Zufallsvariablen $\{Z(x,\omega) : x \in \mathbb{R}^d, \omega \in \Omega\}$, dabei ist
  - $Z(x,\cdot)$ eine **Zufallsvariable** (kurz: $Z(x)$)
  - $Z(\cdot,\omega)$ eine **regionalisierte Variable**, d.h. eine Realisierung der zufälligen Funktion (kurz: $z(x)$)
- Sei $D \subset \mathbb{R}^d$ das **Beobachtungsfenster**, eine konvexe, kompakte Menge; meist ein Rechteck
- Es gebe nun $x_1, \ldots, x_n$ **Messstellen**, $x_i \in D, i = 1, ..., n; \quad z(x_1), ..., z(x_n)$ seien die **Messwerte** an diesen Stellen.

Modellannahmen:

- Das Zufallsfeld $Z(x)$ sei **stationär 2. Ordnung**, d.h.,

$$\mathbb{E}(Z(x)) = \mathbb{E}(Z(x+h)) = m$$

$$\mathbb{E}(Z(x+h) - Z(x))^2 = 2\gamma(h)$$

$$\mathrm{Cov}\,(Z(x), Z(x+h)) = C(h)$$

  $C(h)$ heißt die **Kovarianzfunktion** von $Z(x)$
  $\gamma(h) = \frac{1}{2}\mathbb{E}(Z(x+h) - Z(x))^2$ heißt **Variogramm** von $Z(x)$
  Es gilt:

$$\gamma(h) = C(0) - C(h)$$

- $Z(x) \in \mathcal{C}^1(D)$ für fast alle $\omega \in \Omega$

2 Motivation

Gesucht: Ein Schätzer für die Variogrammfunktion:

$$\gamma(h) := \frac{1}{2}\mathbb{E}(Z(x+h) - Z(x))^2$$

Matheron-Schätzer:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{|x_i - x_j| \approx h} |Z(x_i) - Z(x_j)|^2$$

Die Berechnung erfolgt für $h = h_1, ..., h_{m^d}$ auf einem Gitter im $\mathbb{R}^d$ mit m^d Gitterpunkten.

Ein naiver Ansatz, bei dem alle Paare $Z(x_i), Z(x_j)$ durchsucht werden, erfordert einen Aufwand von $\mathcal{O}(n^2 m^d)$
Mit Hilfe von **binären Suchbäumen** kann dieser Aufwand durch gezielte Suche nach passenden Kombinationen auf $\mathcal{O}(n^2 + m^d(n + \log(n)))$ reduziert werden.

Hier: Wir konstruieren einen Schätzer $\hat{\gamma}$ und schlagen einen Algorithmus zur Berechnung vor mit Aufwand $\mathcal{O}((n + \log(m) + 1)m^d)$

Idee: Konstruiere einen Schätzer $\hat{C}(h)$ für die Kovarianzfunktion $C(h)$ und setze dann

$$\hat{\gamma}(h) := \hat{C}(0) - \hat{C}(h)$$

3 Alternativdarstellung der Kovarianzfunktion

Sei zur Vereinfachung $m = 0$. Dann gilt:

$$\begin{aligned} C(h) &= \mathbb{E}(Z(x)(Z(x+h)) \\ &= \frac{1}{|D \cap (D-h)|} \int_{D\cap(D-h)} \mathbb{E}(Z(x)Z(x+h))\, dx \\ &= \frac{1}{|D \cap (D-h)|}\, \mathbb{E} \int\limits_{D\cap(D-h)} Z(x)Z(x+h)\, dx \\ &= \frac{1}{|D \cap (D-h)|}\, \mathbb{E} \int\limits_{\mathbb{R}^d} Z(-x)\mathbb{1}_D(-x)\; Z(h-x)\mathbb{1}_D(h-x)\, dx \\ &= \frac{1}{|D \cap (D-h)|}\, \mathbb{E}\left((Z(-x)\mathbb{1}_D(-x)) * (Z(x)\mathbb{1}_D(x))\;(h)\right) \end{aligned}$$

Fourier-Transformation

Für $|f| \in L^1(\mathbb{R}^d)$ bezeichne

$$\mathcal{F}\{f(x)\}(s) = \int\limits_{\mathbb{R}^d} e^{-i<s,x>} f(x)\, dx$$

die Fouriertransformierte zu $f(x)$ und

$$\mathcal{F}^{-1}\{f(s)\}(x) = \frac{1}{(2\pi)^d}(p.v.) \int\limits_{\mathbb{R}^d} e^{i<s,x>} f(s)\, ds$$

die Inverse, dann gilt:

$$\mathcal{F}\{f(x) * g(x)\}(s) = \mathcal{F}\{f(x)\}(s)\ \mathcal{F}\{g(x)\}(s)$$

Damit gilt:

$$
\begin{aligned}
C(h) &= c\,\mathbb{E}\left[\mathcal{F}^{-1}\{\mathcal{F}\{Z(x)1\!\!1_D(x)\}\ \mathcal{F}\{Z(-x)1\!\!1_D(-x)\}\}(h)\right] \\
&= \frac{c}{(2\pi)^d}\,\mathbb{E}\int_{\mathbb{R}^d} e^{i<x,s>}\mathcal{F}\{Z(x)1\!\!1_D(x)\}(s)\ \mathcal{F}\{Z(-x)1\!\!1_D(-x)\}(s)\,ds \\
&= \frac{c}{(2\pi)^d}\,\mathbb{E}\int_{\mathbb{R}^d} e^{i<x,s>}\mathcal{F}\{Z(x)1\!\!1_D(x)\}(s)\ \mathcal{F}\{Z(x)1\!\!1_D(x)\}(-s)\,ds
\end{aligned}
$$

mit

$$
c := \frac{1}{|D \cap (D-h)|}
$$

wobei das uneigentliche Integral stets im Sinne des Cauchyschen Hauptwertes **(principal value)** zu verstehen ist.

Um daraus einen Schätzer für $C(h)$ zu konstruieren, sind drei Schritte notwendig:

1. Diskretisierung der Fourier-Transformation
2. Diskretisierung der inversen Fourier-Transformation
3. Vernachlässigen des Erwartungswertes

4 Ein schneller Schätzer der Kovarianzfunktion

Unter den oben gemachten Annahmen hat der Schätzer für $h = (h_1, .., h_d) \in \mathbb{R}^d$ folgende Form:

$$\hat{C}(h) = a(h) \sum_{i,j=1}^{n} b(x_i, x_j, h) Z(x_i) Z(x_j)$$

mit

$$\begin{aligned} a(h) &= \frac{|D|^2 \, (2N)^d \exp(-iN \sum_{j=1}^{d} h_j)}{|D \cap (D-h)| \; n^2 (2\pi)^d m^d} \\ b(x_i, x_j, h) &= \exp(iN \sum_{l=1}^{d} (x_i - x_j)_l) \sum_{m_1=0}^{m-1} \dots \sum_{m_d=0}^{m-1} \exp(-i \frac{2N}{m} \sum_{l=1}^{d} m_l ((x_i - x_j)_l - h_l)) \end{aligned}$$

Für die Herleitung betrachten wir zur Vereinfachung den Fall $d = 2$.
Sei ein äquidistantes Gitter H_m gegeben mit m^2 Gitterpunkten
$h_{ij},\ i,j = 0, ..., m-1$, so dass

$$D \cap (D - h_{ij}) \neq \emptyset \quad \forall h_{ij} \in H_m$$

Sei $S_m \subset [-N, N]^2$ ein äquidistantes Gitter mit m^2 Gitterpunkten
$s_{ij},\ i,j = 0, ..., m-1$.

Ziel: Schätzen von $C(h_{ij})$ für $h_{ij} \in H_m$.

Schritt 1. *Berechne:*

$$\begin{aligned}\tilde{F}(s_{ij}) &:= \frac{|D|}{n}\sum_{k=1}^{n} Z(x_k)e^{-i<x_k,s_{ij}>} \\ &\approx \int_D e^{-i<x,s_{ij}>} Z(x)\,dx \quad i,j=0,...,m-1\end{aligned}$$

Die Berechnung erfolgt direkt mit Aufwand $\mathcal{O}(m^2 n)$.

Schritt 2. *Berechne:*

$$\hat{F}_{ij} := \tilde{F}(s_{ij})\tilde{F}(-s_{ij})$$

Aufwand: $\mathcal{O}(m^2)$

Schritt 3. *Berechne:*

$$\begin{aligned}\tilde{C}(h_{kl}) &:= \frac{c}{(2\pi)^2}\left(\frac{2N}{m}\right)^2 \sum_{i=0}^{m-1}\sum_{j=0}^{m-1} e^{i<s_{ij},h_{kl}>}\hat{F}_{ij} \\ &\approx \frac{c}{(2\pi)^2}\int\limits_{[-N,N]^2} e^{i<s,h_{kl}>}\mathcal{F}\{Z(x)\mathbb{1}_D(x)\}(s)\ \mathcal{F}\{Z(x)\mathbb{1}_D(x)\}(-s)\ ds \\ &\approx C(h_{kl}) \quad k,l = 0,...,m-1\end{aligned}$$

mit

$$c := \frac{1}{|D \cap (D-h)|}$$

Die Berechnung erfolgt nach Transformation des Integrationsbereichs auf $[0,1]^2$ *und des Beobachtungsfensters* D *über Fast-Fourier-Transformation mit Aufwand* $\mathcal{O}(m^2\log(m))$.

Transformation

Im letzten Schritt des Algorithmus berechnen wir

$$\tilde{C}(h_{kl}) := C \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} e^{i<s_{ij},h_{kl}>} \hat{F}_{ij}$$

Für eine Auswertung über Fast-Fourier-Transformation benötigen wir jedoch eine Darstellung der Form

$$\tilde{C}(h_{kl}) := C \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} e^{2\pi i<\tilde{s}_{ij},\tilde{h}_{kl}>} \hat{F}_{ij}$$

mit $\tilde{s}_{ij} = (\frac{i}{m}, \frac{j}{m}) \in [0,1]^2,\ \tilde{h}_{kl} \in \{0, ..., m-1\}^2$

Substitution von s:

Substituiere

$$\tilde{s} = \frac{s}{2N} + \frac{1}{2}$$

Dann ist

$$\begin{aligned} \tilde{C}(h_{kl}) &= C \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} e^{i2N < \tilde{s}_{ij} - \frac{1}{2}, h_{kl} >} \hat{F}_{ij} \\ &= C e^{-iN((h_{kl})_1 + (h_{kl})_2)} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} e^{2\pi i < \tilde{s}_{ij}, \frac{h_{kl} N}{\pi} >} \hat{F}_{ij} \end{aligned}$$

Transformation von h:

Nach unseren Annahmen liegt h_{kl} auf einem äquidistanten Gitter $[-a,a]^2 \subset \mathbb{R}^2$.

Wir benötigen jedoch $\frac{h_{kl}N}{\pi} \in \{0,...,m-1\}^2$

Betrachte deshalb folgende Transformation:

$$\tilde{h} = T(h) := \frac{h+a}{2a}\frac{m\pi}{N}$$

$$\Rightarrow T(h\frac{N}{\pi}) \in \{0,...,m-1\}^d$$

Damit transformieren wir das Beobachtungsfenster D:

$$D \to \tilde{D} = T(D); \quad x_i \to \tilde{x}_i = T(x_i)$$

Dann hat der Schätzer

$$\tilde{C}'(\tilde{h}_{kl}) = Ce^{-iN((\tilde{h}_{kl})_1 + (\tilde{h}_{kl})_2)} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} e^{2\pi i < \tilde{s}_{ij}, \frac{\tilde{h}_{kl} N}{\pi} >} \hat{F'}_{ij}$$

auf $\tilde{D}$ die gewünschte Form.

Beachte: Mit der Transformation von x_i ändert sich auch die Größe $\hat{F}_{ij}$

Aus $\tilde{C}'(\tilde{h}_{kl})$ können wir den gesuchten Schätzer $\tilde{C}(h_{kl})$ durch Rücktransformation gewinnen; es gilt:

$$\tilde{C}'(\tilde{h}_{kl}) = \tilde{C}(T^{-1}(\tilde{h}_{kl}))$$

Bemerkungen:

+ Schnelle Berechnung durch Aufwand $\mathcal{O}(m^2(\log(m) + n + 1))$

- restriktive Modellannahmen: $Z(x) \in \mathcal{C}^1(D)$ $\Rightarrow$ keine Nugget-Effekte modellierbar

- Schätzer nur sinnvoll, falls x_i gleichmäßig verteilt sind auf D $\Rightarrow$ geclusterte Daten müssen bereinigt werden

- Größerer Fehler durch numerische Quadratur

5 Implementierung: Fast-Fourier-Transformation

Ziel: Berechne

$$\tilde{f}_n = \sum_{j=0}^{N-1} e^{-2\pi i n \frac{j}{N}} f_j; \quad n = 0, ..., N-1 \tag{1}$$

$$f_j = \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i j \frac{n}{N}} \tilde{f}_n; \quad j = 0, ..., N-1 \tag{2}$$

wobei die f_j in vielen Fällen die Funktionswerte einer Funktion f an den äquidistanten Stützstellen $v_j = \frac{j}{N}$ sind.
(1) heißt **diskrete Fourier-Transformierte** von f, (2) heißt **Inverse Fourier-Transformierte**.

Sei N eine Zweierpotenz. Dann führt die Auswertung der diskreten Fourier-Transformation auf einen rekursiven Ansatz:
Sei $\omega_N := e^{-\frac{2\pi i}{N}}$ eine N**-te Einheitswurzel**; d.h. $(\omega_N)^N = 1$.
Dann gilt für $n \leq \frac{N}{2} - 1$:

$$
\begin{aligned}
\tilde{f}_n &= \sum_{j=0}^{N-1} e^{-2\pi i n \frac{j}{N}} f_j \\
&= \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i n \frac{2j}{N}} f_{2j} + \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i n \frac{2j+1}{N}} f_{2j+1} \\
&= \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i n \frac{j}{N/2}} f_{2j} + (\omega_N)^n \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i n \frac{j}{N/2}} f_{2j+1}
\end{aligned}
$$

Analog gilt für $n \geq \frac{N}{2}$:

$$
\begin{aligned}
\tilde{f}_n &= \sum_{j=0}^{N-1} e^{-2\pi i n \frac{j}{N}} f_j \\
&= \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i n \frac{j}{N/2}} f_{2j} \;+\; (\omega_N)^n \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i n \frac{j}{N/2}} f_{2j+1} \\
&= \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i \frac{j(n-(N/2))}{N/2}} f_{2j} \;-\; (\omega_N)^{(n-\frac{N}{2})} \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i \frac{j(n-(N/2)}{N/2}} f_{2j+1}
\end{aligned}
$$

Dabei kann man beide Summen wieder durch Fourier-Transformation der Länge $\frac{N}{2}$ berechnen.

Weiter gilt für $N = 1$:

$$\tilde{f}_0 = \sum_{n=0}^{0} e^{-2\pi i \cdot 0} f_j = f_0$$

Dadurch können wir jetzt die diskrete Fourier-Transformation vollständig rekursiv berechnen.

Die Inverse Fourier-Transformation kann man direkt über die normale Fourier-Transformation berechnen; es gilt:

$$\begin{aligned} f_j &= \frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i j \frac{n}{N}}\,\tilde{f}_n \\ &= \frac{1}{N}\overline{\sum_{n=0}^{N-1} e^{-2\pi i j \frac{n}{N}}\,\overline{\tilde{f}_n}} \end{aligned}$$

Beachte: Die FFT wird üblicherweise iterativ implementiert, um Speicher- und Rechenaufwand zu sparen.

Im mehrdimensionalen Falle kann man die n-dimensionale FFT als n-fach iterierte eindimensionale FFT berechnen:

$$\begin{aligned}
\tilde{f}_{(j_1,\dots j_n)} &= \sum_{i_1=0}^{N-1} \cdots \sum_{i_n=0}^{N-1} e^{-2\pi i \frac{1}{N} \sum_{k=1}^{n} i_k j_k} f_{(i_1,\dots,i_n)} \\
&= \sum_{i_1=0}^{N-1} \cdots \sum_{i_{n-1}=0}^{N-1} e^{-2\pi i \frac{1}{N} \sum_{k=1}^{n-1} i_k j_k} \left(\sum_{i_n=0}^{N-1} e^{-2\pi i j_n \frac{i_n}{N}} f_{(i_1,\dots,i_n)} \right)
\end{aligned}$$