WESTFÄLISCHE WILHELMS – UNIVERSITÄT MÜNSTER
Institut für Mathematische Statistik

Ausarbeitung
im Rahmen des Seminars
„Große Abweichungen“

Thema:
Der Satz von Sanov

Vorgelegt von:
Michael Schramm

Inhaltsverzeichnis

1 Einführung

Der Satz von Sanov wurde erstmals von Sanov in den 1930er Jahren für den Spezialfall $\Gamma = \Re$ bewiesen. Er thematisiert die Abweichungen vom Starken Gesetz der Großen Zahlen und befasst sich dabei mit der Konvergenzgeschwindigkeit.

2 Die τ-Topologie

Im Folgenden betrachten wir den Wahrscheinlichkeitsraum $(\mathcal{X}, \mathcal{F})$ und bezeichnen die Menge aller Wahrscheinlichkeitsfunktionen mit $\mathcal{P}$.

Für $P \in \mathcal{P}$, $\mathcal{A} \in \Pi$ und $\epsilon > 0$ sei

$$U(P, \mathcal{A}, \epsilon) = \{P' \in \mathcal{P} : |P'(A_i) - P(A_i)| < \epsilon, i = 1, ..., k\}. \tag{1}$$

Eine Basis der τ-Topologie ist gegeben durch alle Mengen dieser Form. Die τ-Topologie ist die gröbste Topologie, die alle Abbildungen P $\mapsto$ P(F) stetig macht für alle F $\in \mathcal{F}$.

Die Menge aller Partitionen $\mathcal{A} = (A_1, ..., A_k)$ von $\mathcal{X}$ in eine endliche Anzahl von Mengen $A_i \in \mathcal{F}$ wird bezeichnet mit Π.

3 Der Kullback-Leibler-Abstand

Wir betrachten auf einer endlichen Menge $\{1, ..., k\}$ die Wahrscheinlichkeitsverteilungen $P = (p_1, ..., p_k)$ und $Q = (q_1, ..., q_k)$. Der Kullback-Leibler-Abstand (bzw. die relative Entropie) wird definiert als

$$D(P||Q) = \sum_{i=1}^{k} p_i * \log \frac{p_i}{q_i}$$

mit den Konventionen: $0 * \log 0 = 0 * \log \frac{0}{0}$; $t * \log \frac{t}{0} = +\infty$ für $t > 0$.

Der Kullback-Leibler-Abstand der Wahrscheinlichkeitsmaße $P \in \mathcal{P}$, $Q \in \mathcal{P}$ wird definiert als

$$D(P||Q) = \sup_{\mathcal{A} \in \Pi} D(P^{\mathcal{A}}||Q^{\mathcal{A}}) \tag{2}$$

mit $P^{\mathcal{A}} = (P(A_1), \ldots, P(A_k))$, $Q^{\mathcal{A}} = (Q(A_1), \ldots, Q(A_k))$.

4 Empirische Maße

Seien $X_1, \dots, X_n$ Zufallsgrößen mit $X_i \in \mathcal{X}$, dann nennt man das Wahrscheinlichkeitsmaß

$$\widehat{P}_x(F) = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}(F) \ , \ F \in \mathcal{F}$$

das empirische Maß der Folge $X_1, \dots, X_n$. Hierbei bezeichnet δ_x das Diracmaß in x. Für jede Menge A ist $\widehat{P}_x(A)$ der relative Anteil der Treffer der Folge $X_1, \dots, X_n$ in der Menge A.

5 Der Satz von Sanov

Für unabhängige Zufallsgrößen, die bezüglich $Q \in \mathcal{P}$ verteilt sind, erfüllen die resultierenden empirischen Maße das Prinzip großer Abweichungen in der τ-Topologie mit der guten Ratenfunktion $D(\cdot||Q)$. Das bedeutet für $\Gamma \subset \mathcal{P}$:

$$\liminf_{n \to \infty} \frac{1}{n} \log Q^n(\{x : \widehat{P}_x \in \Gamma\}) \geq - \inf_{P \in int_\tau \Gamma} D(P||Q) \tag{3}$$

$$\limsup_{n \to \infty} \frac{1}{n} \log Q^n(\{x : \widehat{P}_x \in \Gamma\}) \leq - \inf_{P \in cl_\tau \Gamma} D(P||Q) \tag{4}$$

und die „Divergenzbälle“

$$B(Q, \alpha) = \{P : D(P||Q) \leq \alpha\} \tag{5}$$

sind kompakt bezüglich der τ-Topologie.

6 Anmerkungen

i. Um sicherzustellen, dass $Q^n(\{x : \widehat{P}_x \in \Gamma\})$ wohldefiniert ist, wird üblicherweise eine Messbarkeitsbedingung an die erlaubten Mengen $\Gamma \subset \mathcal{P}$ gestellt. Alternativ kann $\Gamma \subset \mathcal{P}$ beliebig gewählt werden, wenn $Q^n(\{x : \widehat{P}_x \in \Gamma\})$ als das innere Maß in (3) oder als das äußere Maß in (4) interpretiert wird.

ii. Die Bedingung der unteren Grenze kann verschärft werden, indem die Behauptung für die Topologie τ_0 gezeigt wird. Eine Basis dieser Topologie wird durch Mengen $U_0(\mathcal{P}, \mathcal{A}, \epsilon) = \{P' \in U(P, \mathcal{A}, \epsilon) : P(A_i) = 0 \Rightarrow P'(A_i) = 0\}$ definiert mit $\mathcal{A} = (A_1, \dots, A_k) \in \Pi$ und $\epsilon > 0$

7 Lemma

Sei $\mathcal{P}_n(k)$ die Menge der Wahrscheinlichkeitsverteilungen auf der Menge $\{1,\dots,k\}$ der Form $P = (\frac{n_1}{n},\dots,\frac{n_k}{n}) \in \mathcal{P}_n(k)$ mit den natürlichen Zahlen $n_1,\dots,n_k$. Für ein solches P sei $\mathcal{T}_n(P)$ die Menge aller Tupel der Länge n mit Einträgen aus $\{1,\dots,k\}$, bei denen $i \in \{1,\dots,k\}$ n_i-Mal vorkommt.

i. Es gilt: $|\mathcal{P}_n(k)| \leq (n+1)^k$

ii. Für $P \in \mathcal{P}_n(k)$ und jede Verteilung Q auf $\{1,\dots,k\}$ gilt:
$(n+1)^{-k}e^{-nD(P||Q)} \leq Q^n(\mathcal{T}_n(P)) \leq e^{-nD(P||Q)}$.

8 Beweis Satz von Sanov

8.1 Beweis für (3)

Behauptung: $\liminf_{n\to\infty} \frac{1}{n}\log Q^n(\{x : \widehat{P}_x \in \Gamma\}) \geq -D(P||Q)$ für $P \in int_{\tau_0}(\Gamma)$.

Wähle $P_n \in \mathcal{P}_n(k)$ für alle natürlichen Zahlen n, sodass $P_n \to P^{\mathcal{A}}$ mit $P_n(i) = 0$, wenn $P(A_i) = 0$. Dann gilt: $|P_n(i) - P(A_i)| < \epsilon_n$ für alle $i = 1,\dots,k$ und geeignete $\epsilon_n \to 0$. Und es folgt:

$$Q^n(\{x : \widehat{P}_x \in \Gamma\}) \geq Q^n(\{x : \widehat{P}_x \in U_0(P,\mathcal{A},\epsilon_n)\}) \geq Q^n(\{x : \widehat{P}_x^{\mathcal{A}} = P_n\}$$
$$= (Q^{\mathcal{A}})^n(\mathcal{T}_n(P_n)) \overset{Lemma(ii)}{\geq} (n+1)^{-k}e^{-nD(P_n||Q^{\mathcal{A}})}$$

Durch die Wahl der P_n folgt $D(P_n||Q^{\mathcal{A}}) \to D(P^{\mathcal{A}}||Q^{\mathcal{A}})$ und per Definition $D(P^{\mathcal{A}}||Q^{\mathcal{A}}) \leq D(P||Q)$. Damit gilt die Behauptung.

8.2 Beweis für (5)

Behauptung: $B(Q,\alpha) = \{P : D(P||Q) \leq \alpha\}$ ist kompakt in der τ-Topologie.

Wir betrachten die Menge $\mathcal{M}$ aller endlich additiven Mengenfunktionen auf $(\mathcal{X},\mathcal{F})$ mit Werten im Intervall $[0,1]$. $\mathcal{M}$ ist eine abgeschlossene Menge von $[0,1]^{\mathcal{F}} = \{f : \mathcal{F} \to [0,1]\}$ versehen mit der Produkttopologie, der gröbsten Topologie, die alle Abbildungen $f \mapsto f(F))$ mit $F \in \mathcal{F}$ stetig macht.
Da $[0,1]^{\mathcal{F}}$ nach dem Satz von Tychonov kompakt ist, ist $\mathcal{M}$ als abgeschlossene Teilmenge einer kompakten Menge auch kompakt.
Die Definition des Kullback-Leibler-Abstands überträgt sich unverändert auf P und Q aus $\mathcal{M}$. Offensichtlich ist für jede Partition $\mathcal{A} \in \Pi$ die Teilmenge $\{P \in \mathcal{M} : D(P^{\mathcal{A}}||Q^{\mathcal{A}}) \leq \alpha\} \subset \mathcal{M}$ abgeschlossen und daher kompakt. Als Schnitt für alle $\mathcal{A} \in \Pi$ ist $K = \{P \in \mathcal{M} : D(P||Q) \leq \alpha\}$ auch kompakt. Wenn $Q \in \mathcal{P}$ dann ist die kompakte Menge $K \subset \mathcal{P}$ und daher $K = B(Q,\alpha)$.
Wäre allerdings $P \in \mathcal{M}$ nicht σ-additiv, dann existiert eine absteigende Folge von

Mengen $F_n \in \mathcal{F}$ mit leerem Schnitt und $\lim_{n\to\infty} P(F_n) > 0$, wohingegen aus der σ-Additivität folgt $Q(F_n) \to 0$. Daraus ergibt sich $D(P^{\mathcal{A}_n}||Q^{\mathcal{A}_n}) \to \infty$ für die Partitionen $\mathcal{A}_n = (F_n, F_n^C)$ und daher $D(P||Q) = \infty$ und folglich $P \notin K$. Dies vervollständigt den Beweis, da die Teilraumtopologie $\mathcal{P} \subset \mathcal{M} \subset [0,1]^{\mathcal{F}}$ per Definition der τ-Topologie gleicht.

8.3 Beweis für (4)

Für den Beweis führen wir folgende Notation ein:

$$\Gamma^{\mathcal{A}} = \{P^{\mathcal{A}} : P \in \Gamma\}\ ,\ \Gamma(\mathcal{A}) = \{P \in \mathcal{P} : P^{\mathcal{A}} \in \Gamma^{\mathcal{A}}\} \text{ für } \mathcal{A} \in \Pi$$

1. Behauptung: $cl_\tau\Gamma = \bigcap_{\mathcal{A}\in\Pi} cl_\tau\Gamma(\mathcal{A})$

   Da $\Gamma \subset \Gamma(\mathcal{A})$, bleibt zu zeigen:

   $$P \in cl_\tau\Gamma(\mathcal{A})\ \forall \mathcal{A} \in \Pi \Rightarrow U(P,\mathcal{A},\epsilon) \text{ von P schneidet } \Gamma\, \forall \epsilon > 0.$$

   Sei $\mathcal{A} = (A_1,\ldots,A_k) \in \Pi$ und $P' \in \Gamma(\mathcal{A} \cap U(P,\mathcal{A},\epsilon)$ ($\neq \emptyset$, da $P \in cl_\tau\Gamma(\mathcal{A})$)

   - $P' \in \Gamma(\mathcal{A}) \overset{Def\Gamma(\mathcal{A})}{\Rightarrow} P'(A_i) = \widetilde{P}(A_i), i = 1,\ldots,k$ für ein $\widetilde{P} \in \Gamma$
   - $P' \in U(P,\mathcal{A},\epsilon) \overset{DefU}{\Rightarrow} \widetilde{P} \in U(P,\mathcal{A},\epsilon)$

   Daraus folgt: $U(P,\mathcal{A},\epsilon) \cap \Gamma \neq \emptyset$

2. Behauptung: $\Gamma(\mathcal{A}) \cap B(Q,\alpha) \neq \emptyset$ mit $\mathcal{A} = (A_1 \ldots A_k) \in \Pi$ und $\alpha > \sup_{\mathcal{A}\in\Pi} \inf_{P\in\Gamma} D(P^{\mathcal{A}}||Q^{\mathcal{A}})$

   Aus der Wahl von α folgt die Existenz eines Wahrscheinlichkeitsmaßes $\widetilde{P} \in \Gamma$ mit $D(\widetilde{P}^{\mathcal{A}}||Q^{\mathcal{A}}) < \alpha$. Mit Hilfe dieses $\widetilde{P}$ konstruiere nun das Wahrscheinlichkeitsmaß $P \in \Gamma(\mathcal{A})$ wie folgt:

   $$P(F) = \textstyle\sum_{i=1}^k \frac{\widetilde{P}(A_i)}{Q(A_i)} Q(F \cap A_i)\ ,\ F \in \mathcal{F}$$

   mit der Konvention, dass für $Q(A_i) = 0$ der zugehörige Term gleich Null gesetzt wird. Dann gilt:

   $$P(A_j) = \sum_{i=1}^k \frac{\widetilde{P}(A_i)}{Q(A_i)} Q(A_j \cap A_i) = \frac{\widetilde{P}(A_j)}{Q(A_j)} Q(A_j) = \widetilde{P}(A_j)$$

   Daraus folgt: $P \in \Gamma(\mathcal{A})$, da $P^{\mathcal{A}} = \widetilde{P}^{\mathcal{A}}$ und

   $$D(P^{\mathcal{A}}||Q^{\mathcal{A}}) = \sum_{i=1}^k P(A_i) \log \frac{P(A_i)}{Q(A_i)} = \sum_{i=1}^k \widetilde{P}(A_i) \log \frac{\widetilde{P}(A_i)}{Q(A_i)} = D(\widetilde{P}^{\mathcal{A}}||Q^{\mathcal{A}})$$

   Des weiteren gilt $D(\widetilde{P}^{\mathcal{A}}||Q^{\mathcal{A}}) = D(P^{\mathcal{B}}||Q^{\mathcal{B}})$ für alle Partitionen $\mathcal{B} \in \Pi$, die $\mathcal{A}$ verfeinern. Zusammen mit der Definition von $D(P||Q)$ führt dies zu
   $D(P||Q) = D(\widetilde{P}^{\mathcal{A}}||Q^{\mathcal{A}})$, sodass $P \in B(Q,\alpha)$.

3. Behauptung: $cl_\tau\Gamma \cap B(Q,\alpha) \neq \emptyset$ mit $\mathcal{A} = (A_1 \dots A_k) \in \Pi$
und $\alpha > \sup_{\mathcal{A}\in\Pi} \inf_{P\in\Gamma} D(P^{\mathcal{A}}||Q^{\mathcal{A}})$

   Für jede $A_1, \dots, A_m \in \Pi$ mit $\mathcal{A} \in \Pi$ verfeinert jedes A_i gilt $\Gamma(\mathcal{A}) \subset \Gamma(\mathcal{A}_i)$, $i = 1, \dots, m$. Aus (2.) folgt somit

$$\bigcap_{i=1}^{m} (\Gamma(\mathcal{A}_i) \cap B(Q,\alpha)) \neq \emptyset \tag{6}$$

   Aufgrund der Kompaktheit von $B(Q,\alpha)$ sind auch die Mengen $cl_\tau\Gamma(\mathcal{A}) \cap B(Q,\alpha)$, $\mathcal{A} \in \Pi$ kompakt in der τ-Topologie und alle endlichen Kollektionen dieser Mengen haben nach (6) keinen leeren Schnitt. Es folgt, dass der Schnitt all dieser Mengen ebenfalls nicht leer ist. Mit Hilfe von (1.) folgt die Behauptung.

4. Behauptung: (4)

   Für jede Partition $\mathcal{A} \in \Pi$ gilt:

$$Q^n(\{x : \widehat{P}_x \in \Gamma\}) \leq Q^n(\{x : \widehat{P}_x \in \Gamma(\mathcal{A})\}) = Q^n(\{x : \widehat{P}_x^{\mathcal{A}} \in \Gamma^{\mathcal{A}} \cap \mathcal{P}_n(k)\})$$
$$\stackrel{Lemma(i)}{\leq} (n+1)^k \max_{\overline{P}\in\Gamma^{\mathcal{A}}\cap\mathcal{P}_n(k)} (Q^{\mathcal{A}})^n(\mathcal{T}_n(\overline{P})) \stackrel{Lemma(ii)}{\leq} (n+1)^k e^{-n \inf_{P\in\Gamma} D(P^{\mathcal{A}}||Q^{\mathcal{A}})}$$

   und daher:

$$\limsup_{n\to\infty} \frac{1}{n} \log Q^n(\{x : \widehat{P}_x \in \Gamma\}) \leq \inf_{\mathcal{A}\in\Pi} [-\inf_{P\in\Gamma} D(P^{\mathcal{A}}||Q^{\mathcal{A}})] = -\sup_{\mathcal{A}\in\Pi} \inf_{P\in\Gamma} D(P^{\mathcal{A}}||Q^{\mathcal{A}})$$
$$\stackrel{3.}{\leq} -\sup_{\mathcal{A}\in\Pi} \inf_{P\in cl_\tau\Gamma\cap B(Q,\alpha)} D(P^{\mathcal{A}}||Q^{\mathcal{A}}) \leq -\inf_{P\in cl_\tau\Gamma} D(P||Q)$$

   Damit gilt die Behauptung.

9 Literatur

Imre Csiszár: A simple proof of Sanovs theorem
Bulletin of the Brazilian Mathematical Society, 2006, Volume 37, Number 4, S. 453-459