Konzentrationsungleichungen

Übungsblatt 7

TU Dortmund, Sommersemester 2021

Prof. Dr. Ivan Veselić

Übung 1. *Sei $\mathcal{F}$ die Klasse von lipschitzstetigen Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ mit Lipschitzkonstante 1 bezüglich der 1-Norm, das heißt $|f(x) - f(y)| \leq \sum_{i=1}^n |x_i - y_i|$. Sei $X = (X_1, \ldots, X_n)$ ein Vektor von unabhängigen L^2-Zufallsvariablen. Zeigen Sie*

$$\operatorname{Var}(f(x)) \leq \operatorname{Var}(X_1 + \cdots + X_n) \quad \textit{für alle } f \in \mathcal{F}.$$

Beweis. Nach Efron-Stein gilt:

$$\operatorname{Var}(f(X)) \leq \frac{1}{2}\mathbb{E}\left[(f(X) - f_i'(X))^2\right],$$

wobei $f_i'(X)$ die Zufallsvariable $f(X))$ ist, bei der X_i durch eine unabhängige, identische verteilte Kopie X_i' ersetzt wurde. Wegen der Lipschitzeigenschaft ist

$$|f(X) - f_i'(X)|^2 \leq (X_i - X_i')^2$$

fast sicher. Damit und mit Unabhängigkeit folgt:

$$\begin{aligned}\operatorname{Var}(f(X)) &\leq \frac{1}{2}\sum_{i=1}^n \mathbb{E}[(X_i - X_i')^2] = \frac{1}{2}\sum_{i=1}^n \left[\mathbb{E}[X_i^2] + \mathbb{E}[X_i'^2] - 2\mathbb{E}[X_i]\mathbb{E}[X_i']\right] \\ &= \sum_{i=1}^n \operatorname{Var}(X_i) = \operatorname{Var}(X_1 + \cdots + X_n).\end{aligned}$$

□

Erinnerung: Sei X eine Menge, $\mathcal{A}$ ein Teilmengensystem von X, $F = \{x_1, \ldots, x_n\} \subset X$ eine endliche Menge und $G \subset F$.

- Wir sagen: $\mathcal{A}$ *identifiziert* G, wenn

$$\exists A \in \mathcal{A} \colon A \cap F = G.$$

- Wir definieren

$$s(\mathcal{A}, F) := \#\{G \subset F \colon \mathcal{A} \text{ identifiziert } G\}.$$

  Offensichtlich ist $s(\mathcal{A}, F) \leq 2^n$.

- Der *Vapnik-Chernovenkis-Wachstumskoeffizient* von $\mathcal{A}$ ist

$$s_n(\mathcal{A}) := \sup_{F \in \mathcal{F}_n} s(\mathcal{A}, F),$$

  wobei $\mathcal{F}_n$ die Menge aller n-elementigen Teilmengen von X ist.

- Die *Vapnik-Chernovenkis-Dimension* von $\mathcal{A}$ ist

$$d(\mathcal{A}) := \sup\{n \in \mathbb{N} \colon s_n(\mathcal{A}) = 2^n\}.$$

Übung 2. *a) Sei $X_1 = \mathbb{R}$ und $\mathcal{A}_1$ die Menge aller abgeschlossenen Intervalle, d.h. $\mathcal{A}_1 = \{[a,b]\colon a < b, a, b \in \mathbb{R}\}$. Berechnen Sie $d(\mathcal{A}_1)$.*

b) Sei $X_2 = \mathbb{R}^2$ und $\mathcal{A}_2$ die Menge aller abgeschlossenen Bälle in $\mathbb{R}^2$, d.h. $\mathcal{A}_2 = \{\overline{B}_r(x)\colon r > 0, x \in \mathbb{R}^2\}$, wobei $\overline{B}_r(x) := \{y \in \mathbb{R}^2 \|y - x\|_2^2 \leq r\}$. Berechnen Sie $d(\mathcal{A}_2)$.

Beweis. a) Sei $F = \{x_1, x_2\}$, o.B.d.A. $x_1 < x_2$. Es ist $s(\mathcal{A}_1, F) = 4 = 2^2$, denn $\emptyset$, $\{x_1\}$, $\{x_2\}$, $\{x_1, x_2\}$ werden alle durch Intervalle identifiziert. Das zeigt $s_2(\mathcal{A}_1) = 2^2$.

Sei nun $F = \{x_1, x_2, x_3\}$, o.B.d.A. $x_1 < x_2 < x_3$. Dann wird $\{x_1, x_3\}$ durch kein Intervall in F identifiziert, denn sonst wäre auch x_2 im Intervall. Das zeigt $s_3(\mathcal{A}_1) < 2^3$.

Also folgt $d(\mathcal{A}_1) = 2$.

b) Eine dreiemelentige Menge wird durch Bälle vollständig identifiziert, denn die leere Menge, ein- und zweielementige Mengen und die gesamte Menge wird durch Bälle identifiziert.

Eine vierelementige Menge wird nicht vollständig identifiziert. Um das zu sehen, unterscheiden wir zwei Fälle:

Fall 1: Ein Punkt (z.B. x_4) liegt in der konvexen Hülle der anderen drei. Dann wird $\{x_1, x_2, x_3\}$ niemals durch einen Ball identifziert (denn sonst wäre auch x_4 in dem Ball).

Fall 2: Kein Punkt liegt in der konvexen Hülle der anderen drei. Dann liegen alle Punkte $\{x_1, \ldots, x_4\}$ auf dem Rand der konvexen Hülle und diese ist ein (echtes) Viereck. OE seien die $\{x_1, \ldots, x_4\}$ im Uhrzeigersinn angeordnet.

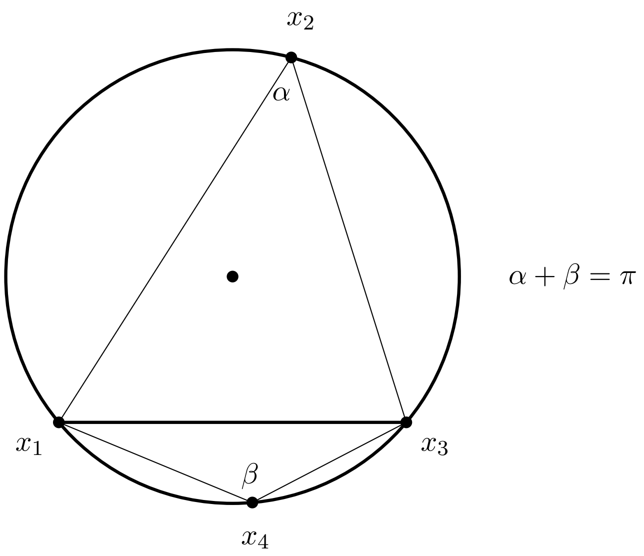
Sei $\overline{B}_s(a)$ der Ball, der $\{x_1, x_3\}$ identifiziert und $\overline{B}_t(b)$ der Ball, der $\{x_2, x_4\}$ identifiziert. Wegen Konvexität des Balles liegt die gesamte Diagonale $[x_1, x_3]$ im Ball, also können wir annehmen (evtl. nach Verkleinerung), dass x_1 und x_3 auf dem Rand von $\overline{B}_s(a)$ liegen und x_2, x_4 außerhalb. Zudem liegen x_2 und x_4 auf unterschiedlichen Seiten der Geraden durch x_1 und x_3.

Wir haben also einen Kreis, der von einer Sehne geschnitten wird.

Nach dem Satz von den Fasskreisbögen, ist für Punkte auf dem Kreisrand zu beiden Seiten der Sehne die Summe der Winkel, die der Sehne gegenüberliegen, gleich π.

Da x_2, x_4 außerhalb liegen, sind die Winkel der Dreiecke (x_1, x_3, x_2) und (x_1, x_4, x_3) bei x_2 und x_4 in Summe kleiner als π.

Führt man die analoge Konstruktion bei x_2, x_4 und $\overline{B}_t(b)$ durch, so findet man, dass das Viereck eine Innenwinkelsumme strikt kleiner als 2π hat - ein Widerspruch.



□

Übung 3. *In der Vorlesung hatten wir bedingte Rademacher-Mittel kennengelernt: Seien $X_{i,j}$ für $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, d\}$ unabhängige Zufallsvariablen mit Werten in $[-1, 1]$ und seien ϵ_i, $i \in \{1, \ldots, n\}$ Rademacher-Zufallsvariablen (gegenseitig und von den X_{ij} unabhängig).*
Wir schreiben $X_i = (X_{i,1}, \ldots, X_{i,d})$ und $X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.
Das bedingte Rademacher-Mittel *war dann definiert worden als*

$$Z = f(X_1, \ldots, X_n) := \mathbb{E}\left[\max_{j=1}^{d} \sum_{k=1}^{n} \epsilon_k X_{k,j} \mid X_1, \ldots, X_n\right].$$

Zeigen Sie, dass Z selbstbeschränkend ist. Genauer gesagt: finden Sie $Z_i = f(X^{(i)})$, so dass

$$0 \le Z - Z_i \le 1 \quad \text{für alle } i = 1, \ldots, n \text{ und } \quad \sum_{i=1}^{n} (Z - Z_i) \le Z.$$

Tipp: Verwenden Sie, wie in der Vorlesung angegeben

$$Z_i := \mathbb{E}\left[\max_{j=1}^{d} \sum_{k=1, k\neq i}^{n} \epsilon_k X_{k,j} \mid X^{(i)}\right].$$

Beweis. Wir haben:

$$\begin{aligned} Z_i &= \mathbb{E}\left[\max_{j=1}^{d} \sum_{k=1, k\neq i}^{n} \epsilon_k X_{k,j} \mid X^{(i)}\right] \\ &= \mathbb{E}\left[\max_{j=1}^{d} \sum_{k=1, k\neq i}^{n} \epsilon_k X_{k,j} \mid X\right] \\ &= \mathbb{E}\left[\sum_{k=1, k\neq i}^{n} \epsilon_k X_{k,j^*} \mid X^{(i)}\right] \end{aligned}$$

für ein j^*, welches von $X^{(i)}$ und $\epsilon^{(i)}$ abhängt. Wir haben:

$$\mathbb{E}[\epsilon_i X_{i,j^*} \mid X] = 0$$

wegen Unabhängigkeit. Damit

$$Z_i = \mathbb{E}\left[\sum_{k=1,k\neq i}^{n} \epsilon_k X_{k,j^*} \mid X\right] \leq \mathbb{E}\left[\max_j \sum_{k=1}^{n} \epsilon_k X_{k,j} \mid X\right] = Z,$$

also $Z - Z_i \geq 0$.

Nun ist

$$Z - Z_i = \mathbb{E}\left[\max_j \sum_k \epsilon_k X_{k,j} - \max_j \sum_{k\neq i} \epsilon_k X_{k,j} \mid X\right]$$

Sein nun j_0 der Index an dem das Maximum der ersten Summe erreicht wird, so gilt

$$... \leq \mathbb{E}\left[\sum_k \epsilon_k X_{k,j_0} - \sum_{k\neq i} \epsilon_k X_{k,j_0} \mid X\right] = \mathbb{E}\left[\epsilon_i X_{i,j_0} \mid X\right].$$

Der Index j_0 ist dabei eine Zufallsvariable, die von ϵ_i abhängig ist, somit sind ϵ_i und X_{k,j_0} nicht unabhängig. Aber $|\epsilon_i|$ und $|X_{k,j_0}| \leq 1$, sodass

$$... \leq 1.$$

Für $\sum_i (Z - Z_i) \leq Z$ berechnen wir

$$\begin{aligned}
\sum_i (Z - Z_i) &= \mathbb{E}\left[\max_j \sum_{k=1}^{n} n\epsilon_k X_{k,j} - \sum_{i=1}^{n} \max_j \sum_{k\neq i} \epsilon_k X_{k,j} \mid X\right] \\
&\leq \mathbb{E}\left[\sum_{k=1}^{n} n\epsilon_k X_{k,j_0} - \sum_{i=1}^{n} \sum_{k\neq i}^{n} \epsilon_k X_{k,j_0} \mid X\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{n} \epsilon_k X_{k,j_0} \mid X\right] = Z.
\end{aligned}$$

(wobei j_0 wieder der Index war, der das erste Maximum maximiert). □

Übung 4. *a) Sei $\mathcal{G}$ die Klasse von Teilmengen $G \subset \mathbb{R}^2$ mit folgender Eigenschaft:*

$$(x, y) \in G, x' \leq x, y' \leq y \Rightarrow (x', y') \in G.$$

Skizzieren Sie zwei sinnvolle (nicht-triviale, also nicht $\emptyset$ oder $\mathbb{R}^2$) Beispiele für Elemente aus $\mathcal{G}$.

b) Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathbb{R}^2$ und unabhängige, $\mathbb{R}^2$-wertige Zufallsvariablen $Z_1, \ldots, Z_n$, die gemäß $\mathbb{P}$ verteilt sind, ist das empirische Maß *P_n definiert als*

$$P_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i}.$$

Finden Sie ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathbb{R}^2$, so dass

$$\lim_{n\to\infty} \mathbb{P}\left[\sup_{G\in\mathcal{G}} |P_n(G) - \mathbb{P}(G)| > \epsilon\right] = 0 \tag{1}$$

gilt.

Beweis. a) Beispiele sind (nach links bzw. unten geöffnete) Halbräume oder die Fläche unter Graphen streng monoton fallender Funktionen.

b) Nach dem Satz von De Hardt-Wright sind $(\mathcal{F}_1, \mathcal{M})$ ein Glivenko-Canteilli-Paar. Hierbei ist

- $\mathcal{F}_M$ die Menge aller $\mathbb{R}^2$ Funktionen, die durch M beschränkt sind und monoton in beiden Koordinaten sind. *Indikatorfunktionen von Mengen $G \in \mathcal{G}$ sind in $\mathcal{F}_1$.*
- $\mathcal{M}$ ist die Menge aller W-Maße μ auf $\mathbb{R}^2$, deren stetiger Anteil μ_C jedem *strikt monotonen Graph* Maß 0 gibt.
- *Strikt monotone Graphen* sind Teilmengen der Form $\{(x, g(x)) \colon x \in \mathbb{R}\}$ für eine strikt monotone Funktion g .
- $\mathcal{F}_1, \mathcal{M}$ ist ein *Glivenko-Cantelli-Paar*, wenn gilt:

  $$\sup_{f \in \mathcal{F}_1} |\int f dP_n - \int f d\mathbb{P}| \to 0 \quad \forall \mathbb{P} \in \mathcal{M}.$$

  (fast sicher, stochastisch oder in Erwartungswert; alles ist äquivalent nach einem Satz, der in der Vorlesung zitiert wurde).

Damit sehen wir, dass das 2-dimensionale Standard-Gaußmaß es tut:

- Es legt gar kein Gewicht auf strikt monotone Graphen
- Die Menge von Indikatorfunktionen von Mengen aus $\mathcal{G}$ ist Teilmenge von Funktionen in $\mathcal{F}_1$.

$\square$

Abgabe bis zum 07.06.2021