Konzentrationsungleichungen

Übungsblatt 11

TU Dortmund, Sommersemester 2021

Prof. Dr. Ivan Veselić

Übung 1. *Seien $X_1, \dots, X_n$ diskrete Zufallsvariablen auf Räumen $\mathcal{X}_1, \dots, \mathcal{X}_n$ mit gemeinsamer Verteilung P und Dichte p und seien $Y_1, \dots, Y_n$ eine weitere Familie von Zufallsvariablen auf den selben Räumen $\mathcal{X}_1, \dots, \mathcal{X}_n$ mit gemeinsamer Verteilung Q und Dichte q. Für $i \in \{1, \dots, n\}$ bezeichnen $P_{1:i}$ und $Q_{1:i}$ die jeweiligen Randdichten der Vektoren $(X_1, \dots, X_i)$ bzw. $(Y_1, \dots, Y_i)$ und bezeichnen $P_{X_i|1\dots i-1}$ und $Q_{Y_i|1\dots i-1}$ die jeweiligen bedingten Dichten von X_i bezüglich $X_1, \dots, X_{i-1}$ unter P bzw. Y_i bezüglich $Y_1, \dots, Y_{i-1}$ unter Q.*

Zeigen Sie, dass für die Kulback-Leibler-Divergenz von P unter Q, definiert durch

$$D(P\|Q) := \sum_{x \in \times_{1\le i\le n} \mathcal{X}_i, p(x)>0} p(x) \ln\left(\frac{p(x)}{q(x)}\right)$$

gilt:

$$D(P\|Q) = \sum_{i=1}^{n} \mathbb{E}_{P_{1:i-1}}\left[D\left(P_{X_i|1\dots i-1}\|Q_{Y_i|1\dots i-1}\right)\right].$$

Beweis. Sei ohne Einschränkung $p(x) > 0$ für alle x. Wir verwenden die (etwas unsaubere) Notation:

$$p(x_1, \dots, x_i) := P(X_1 = x_1, \dots, X_i = x_i)$$

und analog für Q. Dann gilt

$$P_{X_i|1\dots i-1}(X_i = x_i) = \frac{p(x_1, \dots, x_i)}{p(x_1, \dots, x_{i-1})}.$$

Damit folgt:

$$D(P_{X_i|1\dots i-1}\|Q_{Y_i|1\dots i-1}) = \sum_{x_i \in \mathcal{X}_i} \frac{p(x_1, \dots, x_i)}{p(x_1, \dots, x_{i-1})} \ln\left(\frac{p(x_1, \dots, x_i) \cdot q(x_1, \dots, x_{i-1})}{q(x_1, \dots, x_i) \cdot p(x_1, \dots, x_{i-1})}\right).$$

Somit haben wir

$$\begin{aligned}
&\mathbb{E}_{P_{1:i-1}}\left[D(P_{X_i|1\dots i-1}\|Q_{Y_i|1\dots i-1})\right] \\
&= \sum_{x_1,\dots,x_{i-1}} p(x_1, \dots, x_{i-1}) D(P_{X_i|1\dots i-1}\|Q_{Y_i|1\dots i-1}) \\
&= \sum_{x_1,\dots,x_i} p(x_1, \dots, x_i) \ln\left(\frac{p(x_1, \dots, x_i) \cdot q(x_1, \dots, x_{i-1})}{q(x_1, \dots, x_i) \cdot p(x_1, \dots, x_{i-1})}\right).
\end{aligned}$$

Wir zeigen nun die Aussage per Induktion. Der Fall $n = 1$ ist aus letzterer Formel klar.

Gelte die Aussage für $n-1$. Dann ist

$$\begin{aligned}
&D(P\|Q)\\
&=\sum_{x_1,\ldots,x_n} p(x_1,\ldots,x_n)\ln\left(\frac{p(x_1,\ldots,x_n)}{q(x_1,\ldots,x_n)}\right)\\
&=\sum_{x_1,\ldots,x_n} p(x_1,\ldots,x_n)\left[\ln\left(\frac{p(x_1,\ldots,x_n)q(x_1,\ldots,x_{n-1})}{q(x_1,\ldots,x_n)p(x_1,\ldots,x_{n-1})}\right)+\ln\left(\frac{p(x_1,\ldots,x_{n-1})}{q(x_1,\ldots,x_{n-1})}\right)\right]\\
&=\mathbb{E}_{P_{1:n-1}}\left[D(P_{X_n|1\ldots,n-1}\|Q_{Y_n|1\ldots n-1})\right]+\sum_{x_1,\ldots,x_n} p(x_1,\ldots,x_{n-1})\ln\left(\frac{p(x_1,\ldots,x_{n-1})}{q(x_1,\ldots,x_{n-1})}\right)\\
&=\sum_{i=1}^n \mathbb{E}_{P_{1:i-1}}\left[D\left(P_{X_i|1\ldots i-1}\|Q_{Y_i|1\ldots i-1}\right)\right],
\end{aligned}$$

wobei wir im vorletzten Schritt $\sum_{x_n} p(x_1,\ldots,x_n)=p(x_1,\ldots,x_{n-1})$ und im letzten die Induktionshypothese verwendet haben. □

Übung 2. *Sei eine Familie von Graphen $\mathcal{G}(n)$, $n\in\mathbb{N}$ mit endlichen Kantenmengen gegeben. Wir definieren zufällige Teilgraphen $\mathcal{G}(n,p)$ durch Perkolation auf $\mathcal{G}(n)$ mit Parameter $p\in(0,1)$ (Kanten sind unabhängig voneinander aktiviert mit Wahrscheinlichkeit p und mit Wahrscheinlichkeit $1-p$ deaktiviert). Bezeichne $A(n,p)$ das Ereignis, dass $\mathcal{G}(n,p)$ zusammenhängend ist und setze $p_n:=\frac{2\log n}{n}$*

a) Nehmen Sie an, dass gilt:

$$1-\frac{1}{2n}\geq \mathbb{P}\left[A\left(n,p_n\right)\right]\geq 1-\frac{1}{n}.$$

Zeigen Sie dass dann für den totalen Einfluss auf $A(n,p_n)$ folgt:

$$I(A(n,p_n))\geq \frac{1}{8\log^2 n},\quad n\gg 1.$$

b) Nehmen Sie an, dass gilt:

$$3/4\geq \mathbb{P}(A(n,p_n))\geq 1/4.$$

Zeigen Sie dass dann für den totalen Einfluss auf $A(n,p_n)$ folgt:

$$I(A(n,p_n))\geq \frac{n}{32\log^2 n},\quad n\geq 2.$$

Beweis. Das Ereignis $A:=A(n,p_n)$ ist monoton in den einzelnen Bernoulli-Zufallsvariablen, die die Kanten an- und ausschalten. Damit sind wir im Setting von Korollar 4.28 (auch wenn dort n etwas anderes ist, nämlich die Anzahl der Zufallsvariablen, also die Kantenmenge von $\mathcal{G}(n)$). Also gilt:

$$I(A)\geq \frac{\mathbb{P}(A)\log(\frac{1}{\mathbb{P}(A)})}{p_n\log(\frac{1}{p_n})}.$$

Es gilt $p_n\log(\frac{1}{p_n})>0$. Zudem gilt $x\log(\frac{1}{x})\geq x(1-x)$ (Das folgt aus $-\log x\geq 1-x$ bzw. $\log x\leq x-1$).

a) Damit haben wir

$$I(A) \geq \frac{\mathbb{P}(A)(1-\mathbb{P}(A))}{\frac{2\log n}{n}(\log n - \log(2\log n))} \geq n\frac{\mathbb{P}(A)(1-\mathbb{P}(A))}{2\log^2 n} \geq n\frac{(1-\frac{1}{n})\frac{1}{2n}}{2\log^2 n} \geq \frac{1}{8\log^2 n}.$$

Hierbei haben wir verwendet, dass $\log(2\log n) \geq 0$ und $1-\frac{1}{n} \geq \frac{1}{2}$ für hinreichend große n.

Bemerkung: Die Ungleichung $\mathbb{P}(A) \geq 1 - 1/n$ ist bekannt für Erdös-Rény; bei der anderen bin ich mir nicht mal sicher, ob sie in diesem Fall stimmt ...

Interpretation: Wir können mit unserer Methode nicht zeigen, dass der totale Einfluss auf A für $n \to \infty$ nicht gegen 0 geht. Aber immerhin geht er langsam gegen 0.

b) Wir rechnen (analog zu oben)

$$I(A) \geq n\frac{1/16}{2\log^2 n} \geq \frac{n}{32\log^2 n}.$$

□

Übung 3. *Ziel dieser Aufgabe ist, zu zeigen, dass für Wahrscheinlichkeitsverteilungen auf endlichen Mengen die Han-Ungleichung äquivalent zur Subadditivität der Entropie ist. Zur Erinnerung an die Subadditivität der Entropie: Für $Y = f(X_1, \ldots, X_n)$, wobei $X_1, \ldots, X_n$ unabhängige Zufallsvariablen sind, gilt*

$$\mathrm{Ent}(Y) \leq \mathbb{E}\sum_{i=1}^{n} \mathrm{Ent}^{(i)}(Y). \tag{1}$$

Dies war im Fall, dass die Zufallsvariablen endliche Wertebereiche haben, mit Hilfe der Han-Ungleichung bewiesen worden.

Sei nun $\mathcal{X}$ endlich und seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen auf $\mathcal{X}$. Verwenden Sie (1), um daraus die Han-Ungleichung

$$H(X) \leq \frac{1}{n-1}\sum_{i=1}^{n} H\left(X^{(i)}\right)$$

herzuleiten. Dabei gilt für die Verteilung von $X^{(i)} = (X_1, ..., X_{i-1}, X_{i+1}, ...X_n)$, dass

$$\mathbb{P}\left(X^{(i)} = (x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_n)\right) = \sum_{t\in\mathcal{X}} q(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n),$$

wobei $q(x) := \mathbb{P}(X = x)$und $H(X) := -\sum_x q(x)\log q(x)$ ist die Shannon-Entropie von X.

Beweis. Sei Q die Verteilung der Zufallsvariable X und sei P die uniforme Verteilung auf $\mathcal{X}^n$. Sei q die Verteilungsfunktion von X, d.h. $q(x) := \mathbb{P}(X = x)$ und $k := |\mathcal{X}|$. Sei $Y = \mathrm{d}Q/\mathrm{d}P$. Dann gilt $Y(x) = q(x)k^n$ und $\mathbb{E}_P(Y) = 1$. Nun gilt

$$\mathrm{Ent}_P(Y) = D(Q\|P) = \sum q(x)\log(q(x)k^n) = -H(X) + n\log k.$$

sowie

$$\begin{aligned}\mathbb{E}_P\left[\mathrm{Ent}_P^{(1)}(Y)\right] &= \mathrm{Ent}_P(Y) - \sum_{x\in\mathcal{X}^{n-1}}\left(\sum_{t\in\mathcal{X}} q(t,x)\right)\log\left(k^{n-1}\sum_{t\in\mathcal{X}} q(t,x)\right)\\ &= \mathrm{Ent}_P(Y) + H(X^{(1)}) - (n-1)\log(k)\end{aligned}$$

und analog mit 1 ersetzt durch $i \in \{1, \ldots, n\}$. Also haben wir insgesamt:

$$
\begin{aligned}
-H(X) &= -n\log(k) + \mathrm{Ent}_P(Y) \le -n\log(k) + \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Ent}_P^{(i)}(Y)\right] \\
&= -n\log(k) + \sum_{i=1}^{n} \left(\mathrm{Ent}_P(Y) + H(X^{(i)}(Y)) - (n-1)\log(k)\right) \\
&= (-n - n(n-1))\log k + \sum_{i=1}^{n} (-H(X) + n\log k) + H(X^{(i)}(Y))) \\
&= -nH(x) + \sum_{i=1}^{n} H(X^{(i)}(Y))).
\end{aligned}
$$

Umstellen liefert das gewünschte Resultat. □

Übung 4. *Sei μ das Lebesgue-Maß auf $\mathbb{R}$*

a) Seien P und Q Gaußmaße mit Erwartungswerten $\mu, \tilde{\mu} \in \mathbb{R}$ und Varianzen $\sigma, \tilde{\sigma} > 0$. Berechnen Sie die Kullback-Leibler-Divergenz $D(P\|Q)$.

b) Sei P die Gleichverteilung auf dem Intervall $[-1/2, 1/2]$ und Q das Standardgaußmaß auf $\mathbb{R}$ (Erwartungswert 0, Varianz 1). Berechnen Sie die Kullback-Leibler-Divergenz $D(P\|Q)$.

c) Sei nun P die Gleichveteilung auf dem Intervall $[-a, a]$, $a > 0$ und Q das Standardgaußmaß auf $\mathbb{R}$. Finden Sie das a, welches $D(P\|Q)$ minimiert.

Beweis. a) Wir müssen die Abbildungen $X, Y : \mathbb{R} \to [0, \infty)$ aus der Definition identifizieren. Es gilt:

$$
P(A) = \int_A f_P(t)\mathrm{d}t,
$$

also ist $X(t) = f_P(t)$, wobei $f_P(t)$ die Dichte des Gaußmaßes P ist. Analog $Y(t) = f_Q(t)$. Damit ist

$$
\begin{aligned}
D(P\|Q) = \mathbb{E}_P\left[\ln(X/Y)\right] &= \int_{\mathbb{R}} f_P(t)\ln(f_P(t)/f_Q(t))\mathrm{d}t \\
&= \int_{\mathbb{R}} f_P(t)\left(\ln\left(\frac{\tilde{\sigma}}{\sigma}\right) + \frac{-(x-\mu)^2}{2\sigma^2} + \frac{(x-\tilde{\mu})^2}{2\tilde{\sigma}^2}\right)\mathrm{d}t \\
&= \ln\left(\frac{\tilde{\sigma}}{\sigma}\right) - \frac{1}{2} + \int_{\mathbb{R}} f_P(t)\frac{(x-\tilde{\mu})^2}{\tilde{\sigma}^2}\mathrm{d}t \\
&= \ln\left(\frac{\tilde{\sigma}}{\sigma}\right) - \frac{1}{2} + \int_{\mathbb{R}} f_P(t)\frac{(x-\mu)^2 + 2x\tilde{\mu} - \mu^2 + \tilde{\mu}^2}{2\tilde{\sigma}^2}\mathrm{d}t \\
&= \ln\left(\frac{\tilde{\sigma}}{\sigma}\right) - \frac{1}{2} + \frac{\sigma^2}{2\tilde{\sigma}^2} + \frac{(\mu-\tilde{\mu})^2}{2\tilde{\sigma}^2} \\
&= \ln\left(\frac{\tilde{\sigma}}{\sigma}\right) + \frac{1}{2}\left(\frac{\sigma^2}{2\tilde{\sigma}^2} - 1\right) + \frac{(\mu-\tilde{\mu})^2}{2\tilde{\sigma}^2}.
\end{aligned}
$$

b) Wir müssen wieder X und Y identifizieren. Es gilt (analog zu oben): $X(t) = \mathbb{1}_{[-1/2,1/2]}(t)$ und $Y(t) = f(t)$, wobei $f(t)$ die Dichte der Standard-Normalverteilung ist. Es ist

$P(Y = 0) = 0$, also können wir die Kullback-Leibler-Divergenz wie folgt berechnen:

$$\begin{aligned} D(P\|Q) &= \mathbb{E}_P\left[\ln\left(\frac{X}{Y}\right)\right] = \int_{-1/2}^{1/2} \ln\left(f(x)^{-1}\right) \mathrm{d}x = \int_{-1/2}^{1/2} \ln(\sqrt{2\pi}) + \frac{x^2}{2} \mathrm{d}x \\ &= \frac{1}{2}\ln(2\pi) + \frac{2(1/2)^3}{6} = \frac{1}{2}\ln(2\pi) + \frac{1}{24}. \end{aligned}$$

c) Jetzt ist $X(t) = \frac{1}{2a}\mathbb{1}_{[-a,a]}(t)$. Wir berechnen analog zu Teil b):

$$D(P\|Q) = \ln(\sqrt{2\pi}) - \ln(2a) + \frac{a^2}{6}.$$

Kurvendiskussion (oder Wolframalpha) zeigt, dass das minimal wird bei

$$a = \sqrt{3}.$$

$\square$

Abgabe bis zum 05.07.2021