Informationstheorie, FS 2017 Dr. Luis Haug
ETH Zürich, Dept. Informatik Nico Gorbach

Lösungen zu Übungsblatt 5

5.1

Siehe Lösung zu Aufgabe 4.15 in [MacKay, S. 87].

5.2

a) Wegen der Unabhängigkeit der (X_i, Y_i) gilt

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \frac{P(X^n)P(Y^n)}{P(X^n, Y^n)} &= \lim_{n\to\infty} \frac{1}{n} \log \prod_{i=1}^{n} \frac{P(X_i)P(Y_i)}{P(X_i, Y_i)} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log \frac{P(X_i)P(Y_i)}{P(X_i, Y_i)} \\
&= E\Big(\log \frac{P(X)P(Y)}{P(X,Y)} \Big) \\
&= D_{KL}(P(X,Y) \| P(X)P(Y)) \\
&= I(X,Y)
\end{aligned}$$

unter Verwendung des Gesetzes der grossen Zahlen.

b) Wir können

$$P(X_1, \ldots, X_n) = 2^{\log P(X_1,\ldots,X_n)} = 2^{\sum_{i=1}^{n} \log P(X_i)}$$

schreiben, wobei die zweite Umformung wegen der Unabhängigkeit der X_i gilt. Damit folgt

$$\begin{aligned}
\lim_{n\to\infty} (P(X_1, X_2, \ldots, X_n))^{\frac{1}{n}} &= \lim_{n\to\infty} 2^{\frac{1}{n}\sum_{i=1}^{n} \log P(X_i)} \\
&= 2^{-H(X)},
\end{aligned}$$

denn $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log P(X_i) = -H(X)$ aufgrund des Gesetzes der grossen Zahlen, und weil wir die Grenzwertbildung und die Anwendung der Funktion $x \mapsto 2^x$ vertauschen dürfen, weil diese Funktion stetig ist ([**TODO**Check]).

5.3

a) $H(X) = -0.6 \log 0.6 - 0.4 \log 0.4 = 0.97095$ bits.

b) Die typische Menge $T_{n,\varepsilon} = \{\mathbf{x} \in \{0,1\}^n : H(X) - \varepsilon \leq -\frac{1}{n} P(\mathbf{x}) \leq H(X) + \varepsilon\}$ enthält für $n = 25$ und $\varepsilon = 0.1$ genau diejenigen $\mathbf{x} \in \{0,1\}^{25}$ mit $-\frac{1}{n} \log P(\mathbf{x}) \in (0.87095, 1.07095)$; aus der Tabelle kann man ablesen, dass das genau die Folgen mit $11 \leq k \leq 19$ sind, wobei k die Anzahl der Einsen in der Folge bezeichnet (in der Tabelle ist $p = P(1) = 0.6$).

k	$\binom{n}{k}$	$\binom{n}{k}p^k(1-p)^{n-k}$	$F(k)$	$-\frac{1}{n}\log P(\mathbf{x})$
0	1	0.000000	0.000000	1.321928
1	25	0.000000	0.000000	1.298530
2	300	0.000000	0.000000	1.275131
3	2300	0.000001	0.000001	1.251733
4	12650	0.000007	0.000008	1.228334
5	53130	0.000045	0.000054	1.204936
6	177100	0.000227	0.000281	1.181537
7	480700	0.000925	0.001205	1.158139
8	1081575	0.003121	0.004326	1.134740
9	2042975	0.008843	0.013169	1.111342
10	3268760	0.021222	0.034392	1.087943
11	4457400	0.043410	0.077801	1.064545
12	5200300	0.075967	0.153768	1.041146
13	5200300	0.113950	0.267718	1.017748
14	4457400	0.146507	0.414225	0.994349
15	3268760	0.161158	0.575383	0.970951
16	2042975	0.151086	0.726469	0.947552
17	1081575	0.119980	0.846448	0.924154
18	480700	0.079986	0.926435	0.900755
19	177100	0.044203	0.970638	0.877357
20	53130	0.019891	0.990529	0.853958
21	12650	0.007104	0.997633	0.830560
22	2300	0.001937	0.999571	0.807161
23	300	0.000379	0.999950	0.783763
24	25	0.000047	0.999997	0.760364
25	1	0.000003	1.000000	0.736966

$P(T_{n,\varepsilon})$ ist damit gegeben durch $F(19)-F(10) = 0.97068-0.034392 = 0.936246$, wobei F die kumulative Verteilungsfunktion der Zufallsvariablen $\mathbf{x} \mapsto k$ bezeichnet (die eine Folge auf die Anzahl der in ihre enthaltenen Einsen abbildet). Aus der zweiten Spalte lässt sich die Kardinalität ermitteln:

$$|T_{n,\varepsilon}| = \sum_{k=11}^{19} \binom{25}{k} = 26366510.$$

c) $S_{n,\delta}$ ist nach Definition eine Menge kleinstmöglicher Kardinalität, die $P(S_{n,\delta}) > \delta$ erfüllt. Um eine solche Menge zu finden, definieren wir mit eine Folge von Teilmengen $\Omega_i \subset \{0,1\}^{25}$ durch $\Omega_0 := \emptyset$ und $\Omega_i := \Omega_{i-1} \cup \{\hat{\mathbf{x}}_i\}$, wobei $\hat{\mathbf{x}}_i$ ein Element mit grösster Wahrscheinlichkeit aus dem Komplement von Ω_{i-1} ist; dann können wir $S_{n,\delta} = \Omega_{i_\delta}$ setzen, wobei i_δ der kleinste Index ist so dass $P(\Omega_i) \geq 1-\delta$ gilt. Da in unserem Beispiel die Wahrscheinlichkeit von Elementarereignissen monoton mit k wächst, beginnen wir mit den Elementen mit maximalem k. Aus der Tabelle kann man ablesen, dass $P(k \geq 12) > 0.9 > P(k \geq 13)$ (denn $P(k \geq 12) = 1 - F(11)$ etc.), und daher enthält das so konstruierte $S_{n,\delta}$ zunächst alle Folgen mit $k \geq 13$; es gilt $P(k \geq 13) = 1 - F(12) = 0.846232$ und $|\{\mathbf{x} \,|\, k \geq 13\}| = 16777216$. Die Anzahl der

Folgen in $S_{n,\delta}$ mit $k = 12$ ist

$$\left\lceil \frac{0.9 - P(k \geq 13)}{p^{13}(1-p)^{12}} \right\rceil = 3680691.$$

Damit ist $|S_{n,\delta}| = 16777216 + 3680691 = 20457907$.

d) Der Schnitt $T_{n,\varepsilon} \cap S_{n,\delta}$ enthält alle Folgen mit $13 \leq k \leq 19$ und 3680691 Folgen mit $k = 12$. Damit gelten $P(T_{n,\varepsilon} \cap S_{n,\delta}) = 0.870638$ und $|T_{n,\varepsilon} \cap S_{n,\delta}| = 16708810 + 3680691 = 20389501$.