4. Übung Informationstheorie SS18

1. Kartenmischen I: "Überhand": ein Kartenstapel mit N Karten wird gemischt, indem pro Durchgang der Stapel an k ($\ll N$) zufällig gewählten Stellen geteilt und die Teile in umgekehrter Reihenfolge übereinandergelegt wwrden. Wie viele Durchgänge sind mindestens erforderlich, um eine "gute" Mischung zu erreichen (also eine annähernde Gleichverteilung auf der Menge der Permutationen - vergleichen Sie die Entropie dieser Verteilungmit der gemeinsamen Entropie der einzelnen Mischvorgänge)?

2. Kartenmischen II: "Riffeln": diesmal wird der Stapel geteilt und die beiden Teile ineinander "gefächert". Ein solcher Mischvorgang kann durch die Menge der Karten im gemischten Stapel beschrieben werden, die aus dem oberen Teil stammen (allerdings entsprechen nicht alle solchen Teilmengen unterschiedlichen Permutationen). Wieder: wie oft muss man mindestens mischen?

3. Kartenmischen III: es wird jeweils die oberste Karte an eine zufällig gewählte Stelle gesteckt.

4. Der "Bibelcode": in einem Text von N Buchstaben $x_1 \dots x_N$ werden (für fixes l) die Folgen $x_i x_{i+d} \dots x_{i+(l-1)d}$ mit $d > 1, 1 \leq i \leq N - (l-1)d$ betrachtet. Wenn eine solche Folge ein bestimmtes (oder auch nur irgendein mehr oder weniger sinnvolles) Wort ergeben, freuen wir uns.

   (a) Wie viele solche Folgen gibt es?

   (b) Nehmen Sie an, dass die Buchstaben unabhängig und gleichverteilt aus dem Alphabet M mit $|M| = m$ gezogen wurden. Wie groß ist die Wahrscheinichkeit, dass eine bestimmte Folge mit einem vorgegebenen Wort $y_1 \dots y_l$ übereinstimmt? Wie groß ist der Erwartungswert der Anzahl aller solcher Folgen? Rechnen Sie speziell für $N = 2^{22}$, $m = 27$, $l = 7, 8, 9$.

5. Das vorige Beispiel wird natürlich einer natürlichen Sprache nicht gerecht. Nehmen Sie an, dass der Text aus einer Quelle kommt, die eine Entropie hat. H_1 sei die Entropie der Verteilung der einzelnen Buchstaben, $H_l < H_1$ die Entropie pro Buchstabe der "sinnvollen" Wörter der Länge l. Wir können annehmen, dass die Buchstaben in unseren Teilfolgen unabhängig gezogen werden (zumindest, wenn d nicht zu klein ist). Nach Shannon-McMillan gibt es $\approx 2^{lH_l}$ "typische" Wörter der Länge l, für die die Wahrscheinlichkeit, sie in einer Teilfolge zu finden 2^{-lH_1} beträgt. Bestimmen Sie den Erwartungswert der Anzahl der "Funde" für ein solches Wort und für alle solchen Wörter. Setzen Sie im Zahlenbeispiel $H_1 = 4.2$ und $H_l = 1.6$.