Theoretische Grundlagen der Informatik

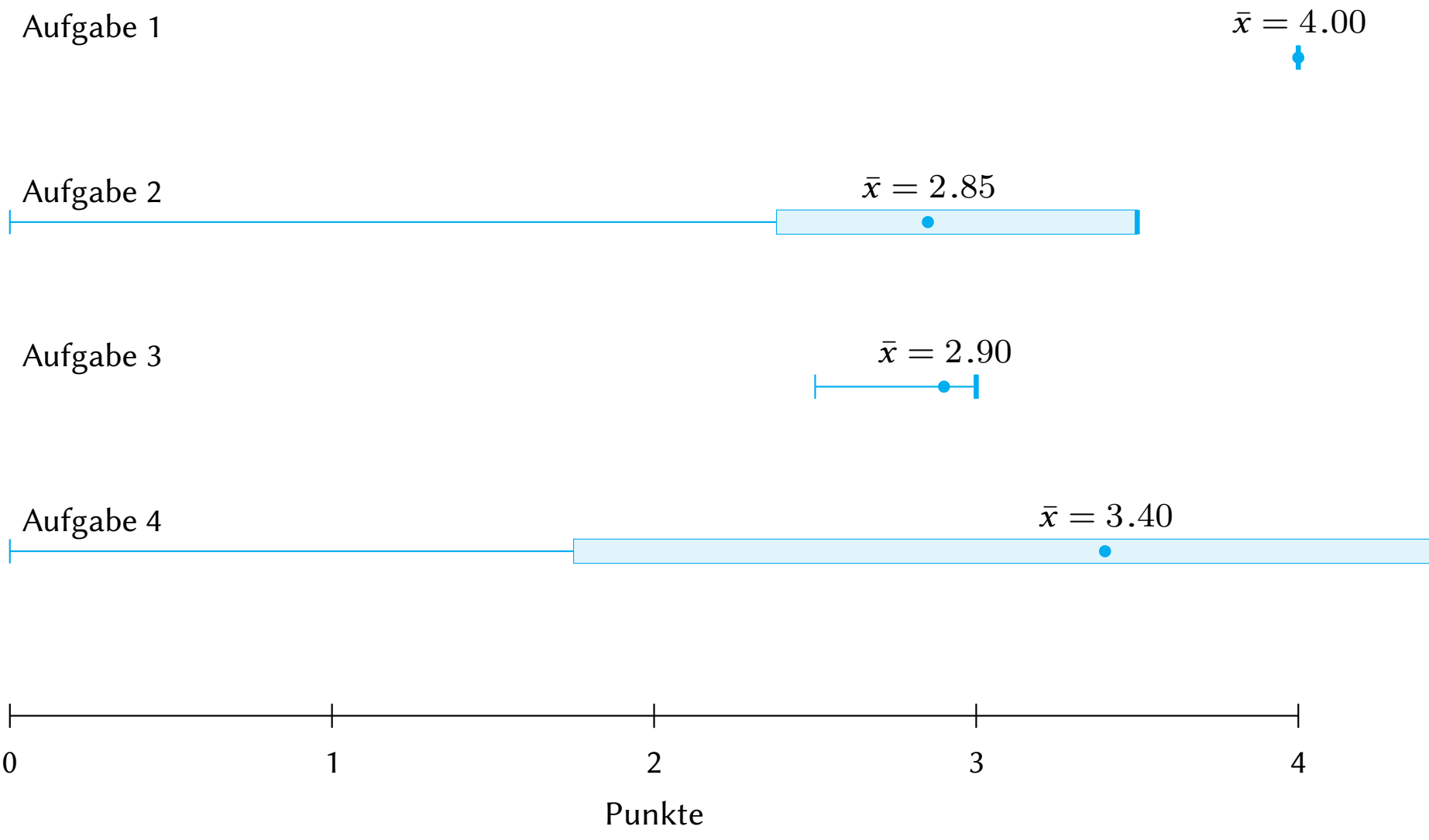
Tutorium 11 — Wintersemester 2013/14

Moritz Klammler — 28. Januar 2014

Organisatorisches

- Es wird (inklusive heute) noch drei Tutorien geben.
- Es wird insgesamt sieben benotete und ein unbenotetes Übungsblatt geben.
- Klausuranmeldung ist angeblich möglich.

6. Übungsblatt (10 gültige Abgaben)



Anmerkungen zum 6. Übungsblatt

- Übungsblätter bitte sorgfältig aufbewahren (Beweismittel für Euch).
- Um zu zeigen, dass $L \in \mathcal{NP}$, genügt es, einen polynomiellen *Verifizierer* anzugeben. Ein Suchalgorithmus ist nicht erforderlich.
- Bei Reduktionen immer eine Transformation von einer Instanz des „schwierigen“ Problems in eine Instanz des „neuen“ Problems angeben. (Die Gegenrichtung ist klar.)

Tagesthemen

- Komprimierbarkeit (Kolmogorow-Komplexität)
- Entropie
- Codierungen

```
#include <stdio.h>
#include <stdlib.h>
#include <time.h>

#define CC_FIRST ' '
#define CC_LAST  '~'

void pwgen(char * buffer, size_t length)
{
  const unsigned int seed = time(NULL);
  size_t i;
  srand(seed);
  for (i = 0; i < length; i++)
    {
      buffer[i] = CC_FIRST + rand() % (CC_LAST - CC_FIRST);
    }
}
```

Kolmogorow-Komplexität I

Sei Σ ein Alphabet und $w \in \Sigma^*$. Wir definieren die Sprache

$$C(w) = \{\langle M \rangle x \in \Sigma^* : M(x) = w\}$$

aller Gödelnummern $\langle M \rangle$ mit hardgecodeter Eingabe x, die – angesetzt auf ein leeres Band – w auf das Band schreiben und halten.

Die Länge des kürzesten Worts aus $C(w)$

$$K(w) = \min\{|c| : c \in C(w)\}$$

heißt **Kolmogorow-Komplexität** von w.

Kolmogorow-Komplexität II

Für jedes $n \in \mathbb{N}_0$ existiert ein[1] $w \in \Sigma^n$ mit $K(w) \geq n$.

Die Funktion $K : \Sigma^* \to \mathbb{N}_0$ ist nicht berechenbar.

[1] sehr sehr viele

```
$ dd if=/dev/zero    of=zero.dat   bs=1K count=10K
10240+0 records in
10240+0 records out
10485760 bytes (10 MB) copied, 0.0194976 s, 538 MB/s

$ dd if=/dev/urandom of=random.dat bs=1K count=10K
10240+0 records in
10240+0 records out
10485760 bytes (10 MB) copied, 0.940868 s, 11.1 MB/s

$ gzip -k zero.dat random.dat

$ ls -lh
-rw-r----- 1 tux tux 10M Jan 28 16:07 zero.dat
-rw-r----- 1 tux tux 10K Jan 28 16:07 zero.dat.gz
-rw-r----- 1 tux tux 10M Jan 28 16:07 random.dat
-rw-r----- 1 tux tux 11M Jan 28 16:07 random.dat.gz
```

Information und Entropie

Sei $\mathcal{U}$ ein endliches Universum und X eine Zufallsvariable aus $\mathcal{U}$ mit Verteilung p.

Die **Information** einer Realisierung x von X ist definiert als

$$I(x) = -\log\Big(p(x)\Big) \quad .$$

Die **Entropie** der Quelle X ist definiert als der Erwartungswert der Information.

$$H(X) = -\sum_{x \in \mathcal{U}} p(x) \log\Big(p(x)\Big)$$

Codierungen

- Identität (ein Code-Symbol pro Zeichen)
- unär mit Trennzeichen
- Wörter fixer Länge (ASCII, n-bit Integer, IEEE-754, ...)
- **Präfixcodes (Huffman-Code)**

Idee: Für Quelle X mit Verteilung p, finde Codierung C, sodass der Erwartungswert der Wortlänge

$$\sum_{x \in \mathcal{U}} p(x)|C(x)|$$

minimiert wird.

Beispiele

- we were arrested after dad ate deer eggs
- the quick brown fox jumps over the sleazy dog