Ergänzungen zum maschinellen Übersetzen natürlicher Sprachen

3. Übungsblatt

2016-05-10

Exercise 1

Prove Gibb's inequality: for every $p, q \in \mathcal{M}(A)$ we have

$$\sum_a p(a) \cdot \log p(a) \geq \sum_a p(a) \cdot \log q(a) \ .$$

Hint: for each $x \geq 0$ it holds that $\log x \leq x - 1$.

Exercise 2

By throwing a die $m \geq 1$ times you win a prize, if you obtain every number $i \in \{1, \ldots, 6\}$ exactly m_i times, where $m = m_1 + \ldots + m_6$. Assume you can produce your own die subject to the following condition: opposing sides must have the same probability. Define the underlying probability model. Which probability distribution would you chose?

Exercise 3

Let $n \in \mathbb{N}$. Let $\Omega_1, \ldots, \Omega_n$ and $X_1, \ldots, X_n$ be non-empty, finite sets. For every $i \in \{1, \ldots, n\}$, let $p^{(i)} \colon \Omega_i \to \mathcal{M}(X_i)$ be a Ω_i-probability model. Let $\Omega = \Omega_1 \times \ldots \times \Omega_n$ and $X = X_1 \times \ldots \times X_n$. We define the Ω-probability model $p \colon \Omega \to \mathcal{M}(X)$ by

$$p_{(\omega_1,\ldots,\omega_n)}(x_1, \ldots, x_n) = p^{(1)}_{\omega_1}(x_1) \cdot \ldots \cdot p^{(n)}_{\omega_n}(x_n)$$

for every $\omega_i \in \Omega_i$, $x_i \in X_i$, $i \in \{1, \ldots, n\}$.

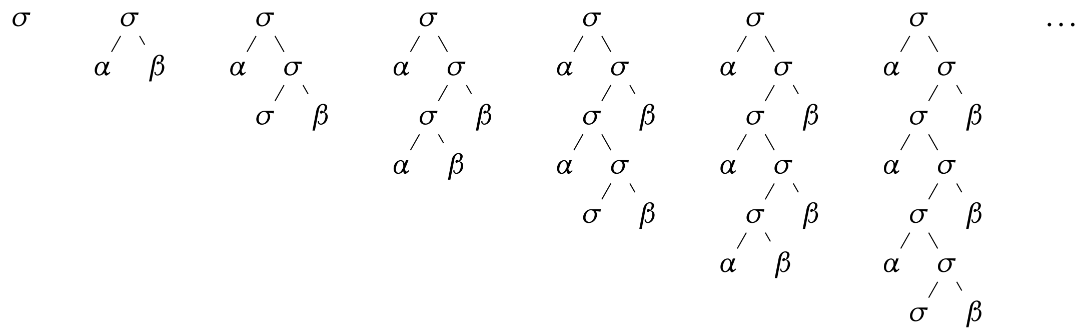
Let c be an X-corpus. For every $i \in \{1, \ldots, n\}$, let c_i be an X_i-corpus such that $c_i(x') = \sum_{x=(x_1,\ldots,x_n)\in X \colon x_i = x'} c(x)$. Show that $\mathrm{mle}_p(c) = \mathrm{mle}_{p^{(1)}}(c_1) \times \ldots \times \mathrm{mle}_{p^{(n)}}(c_n)$.

Exercise 4

Let Σ be an alphabet, V a set, $k \in \mathbb{N}$, and $t, t', t'_1, \ldots, t'_k \in U_\Sigma(V)$. Formally define the following notions from the lecture:

- pos,
- $t(w)$ for $w \in \mathrm{pos}(t)$,
- $t|_w$ for $w \in \mathrm{pos}(t)$,
- $t[t']_w$ for $w \in \mathrm{pos}(t)$,
- $c[t'_1, \ldots, t'_k]$ for $k \in \mathbb{N}$ and $c \in C^k_\Sigma$.

Exercise 5
Construct a regular tree grammar G, such that $[\![G]\!]$ contains exactly the trees of the following form.



Note that every second tree has a σ leaf. Give some elements of $D^q(G)$ for some states q of G. Is your G unambiguous? Is it deterministic? If not, can you find a deterministic grammar?