

**Maschinelles Lernen und Data Mining**  
Sommersemester 2013  
**Übungsblatt 5**

*Besprechung des Übungsblattes am 13.06.2013*

**Aufgabe 5-1**      Frequentistische Statistik vs. Bayessche Statistik

Alice und Bob spielen ein Spiel bei dem derjenige gewinnt, der zuerst 6 Punkte erreicht. Dazu wird zu Beginn des Spiels in einem Raum (den Alice und Bob nicht einsehen können) eine Kugel auf einem Billardtisch gerollt. Die Kugel bleibt (gleichverteilt) auf einer komplett zufälligen Position liegen die anschließend markiert wird. Danach werden Punkte unter Zuhilfenahme einer zweiten Kugel entschieden, die ebenfalls zufällig über den Tisch gerollt wird. Landet die zweite Kugel links der Markierung bekommt Alice den Punkt, landet die Kugel rechts der Markierung bekommt Bob den Punkt. Keiner der beiden hat einen Einblick auf den Billardtisch, Ihnen werden jedoch die Punktestände mitgeteilt. Welche Chance hat Alice zu gewinnen, wenn es bereits 5:3 steht?

- (a) Schreiben Sie ein Programm um diese Wahrscheinlichkeit zu schätzen.
- (b) Überlegen Sie wie diese Wahrscheinlichkeit mittels frequentistischer Statistik bzw. mit bayesianischer Statistik berechnet werden kann.

**Aufgabe 5-2**      Personengröße

Angenommen die Größe einer Population  $\subset \mathbb{R}$  sei normalverteilt:

$$P_{\mathbf{w}}(\mathbf{x}_i) = \mathcal{N}(\mathbf{x}_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{x}_i - \mu)^2}{2\sigma^2}\right)$$

Berechnen Sie die entsprechenden ML Schätzer für die vier Größendatensätze unter `body_sizes.txt` und visualisieren Sie die dazugehörigen Verteilungen. Ist der Schätzer hilfreich für das Verständnis der Daten?