Prof. Dr. Johannes Blömer

28.01.2016
submission due: 4.02.2016, F1.110, 11:00

Clustering Algorithms

WS 2015/2016

Handout 11

Exercise 1:
Denote the center of a set $A \subset \mathbb{R}$ by $\mu(A) = \frac{1}{|A|} \sum_{a \in A} a$. Given a set $P \subset \mathbb{R}$, we draw n points uniformly at random from P. Denote by x_i the i-th point that is drawn uniformly at random from P, and let $X = \{x_1, \dots, x_n\}$.
In Handout 5, we observed that

$$E[\mu(X)] = E[x_i] = \mu(P) \text{ and } \mathrm{Var}(\mu(X)) = \frac{1}{n}\mathrm{Var}(x_i).$$

Now we want to estimate the variance $\mathrm{Var}(x_i)$ by

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu(X))^2.$$

Show that this estimate is biased, i.e.

$$E\left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu(X))^2\right] = \frac{n-1}{n}\mathrm{Var}(x_i).$$