The two-sample Kolmogorov-Smirnov Test

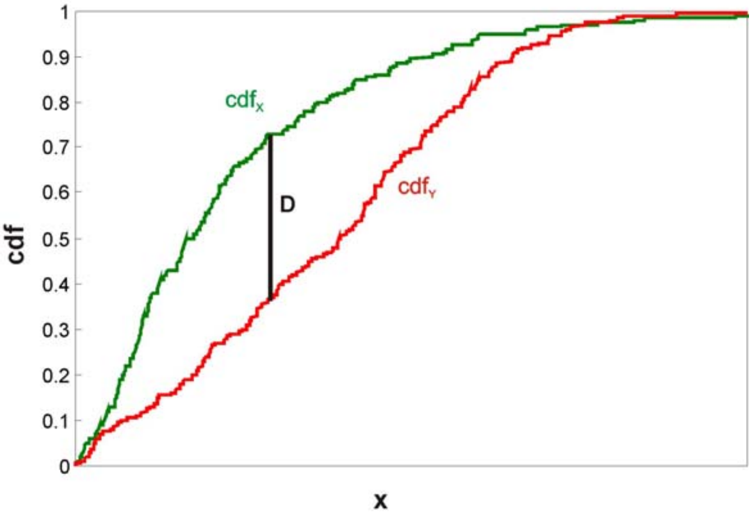
We used a two-sample Kolmogorov-Smirnov test to quantitatively compare the virtual test data set X_i with simulated probe data-sets Y_j. The principle of the test can be found in text books on statistics (e.g. (1, 2)), and shall be briefly sketched here.
Let us assume two samples X_i and Y_j with size n and m drawn from continuous distributions. The two samples can be characterized by their empirical cumulative density functions

$$cdf_X(x) = \frac{\#i : X_i \le x}{n} \text{ and } cdf_Y(x) = \frac{\#i : Y_i \le x}{m}$$

which are defined as the proportion of observed values that are less than or equal to x. The test is based on finding the maximum distance $D = \max(|cdf_X(x) - cdf_Y(x)|)$. Suppose that both X_i and Y_j are drawn from the same distribution, and their values are such that D=d. Since a large value of D would appear to be inconsistent with the null hypothesis that both samples are drawn from the same distribution, it follows that the P-value for this data-set is given by

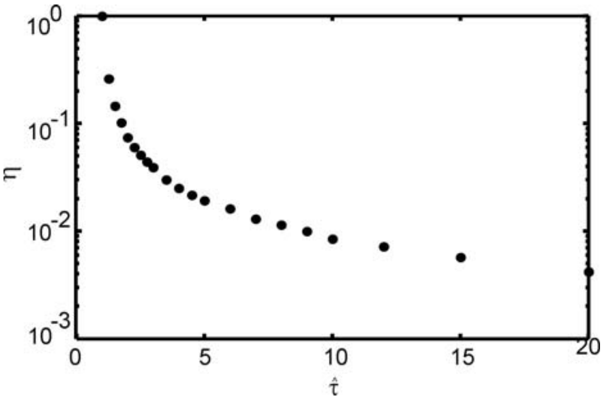
$$P - value = prob(D \ge d)$$

Note that prob specifies the probability for $D \ge d$ under the assumption that H_0 is correct. The key argument for the versatility of the Kolmogorov-Smirnov test is the proposition that $prob(D \ge d)$ is the same for any continuous distribution. The distribution of D as function of n and m can be found in tables (2) and is implemented for calculation of the P-value in the Matlab function kstest2 (The MathWorks, Natrick, MA).



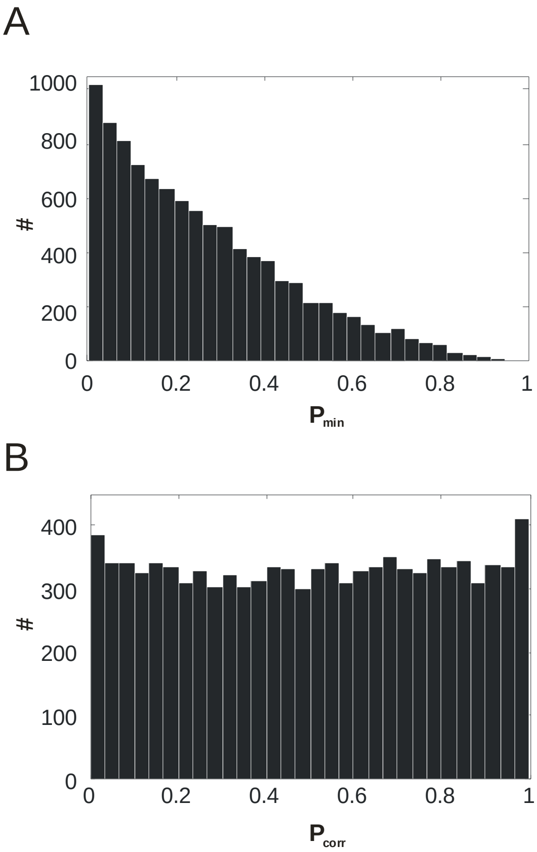
1. Ross, S. M. Introduction to probability and statistics for engineers and scientists, 3rd edition, Elsevier, San Diego
2. Conover, W. J.. Practical nonparametric statistics, 3rd edition, Wiley, New York

Supplemental Figure 1



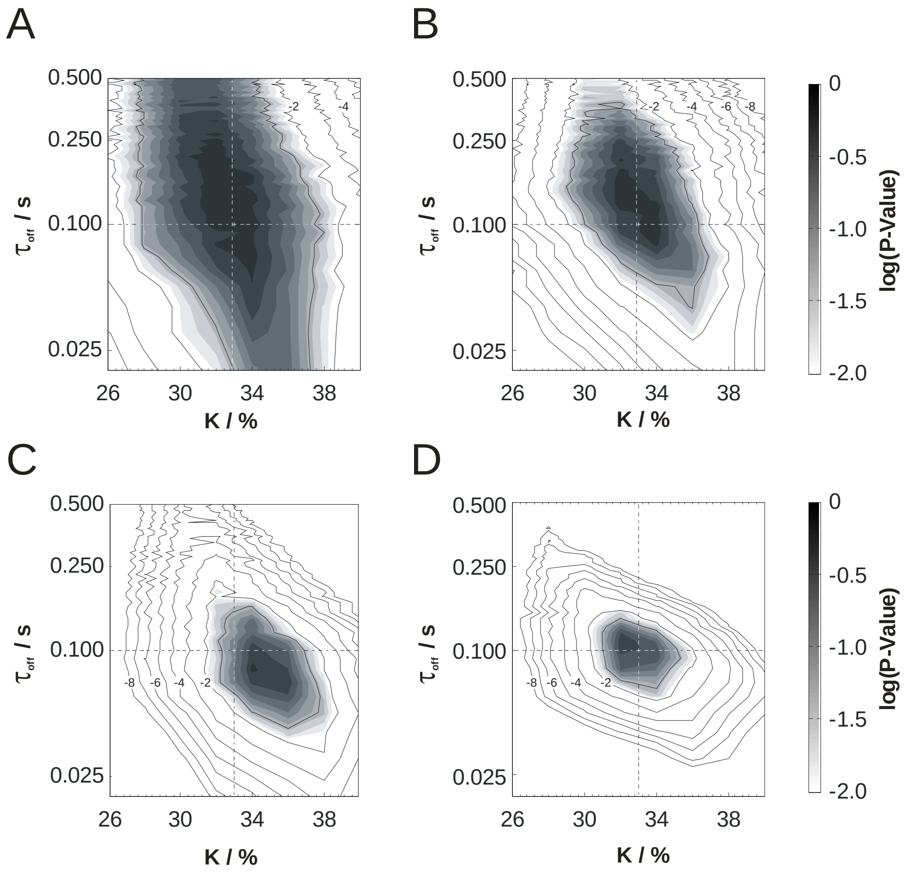
Supplemental Figure 1. Escape-probability η as a function of $\hat{\tau}$. Data were obtained from Monte Carlo simulations of hop diffusion, by determining $\hat{\tau}$ for various η.

Supplemental Figure 2



Supplemental Figure 2: Correction of the P-value distribution. We simulated free diffusion according to Fig. 3 and calculated the minimum P_{min} of the P-values obtained for n=1, n=2 and n=3. The distribution is clearly non-uniform (**A**). Using the transformation $P_{corr} = 1-(1-P_{\min})^{n_{\max}}$, a uniform distribution of P_{corr} is generated, which allows for interpretation of P_{corr} as P-value (**B**).

Supplemental Figure 3



Supplemental Figure 3: P-values as function of K and τ_{off} for various ratios D_A/D_{AB}. Tests were performed on the data-sets described in Fig. 12C. The panels show results for $D_A/D_{AB}=2$ (**A**), 3 (**B**), 4 (**C**) and 10 (**D**). Interestingly, already a ratio of three is sufficient to significantly restrict the parameter range for τ_{off}. Data obtained for a ratio of 2, however, contain hardly any information on the interaction lifetime.