Problem: Estimate expected overlaps A_B and/or B_A if we only know summary statistics of both samples, such as diversity N: $\{N(A), n_i(A), c_i(A), N(B), n_i(B), c_i(B)\}$
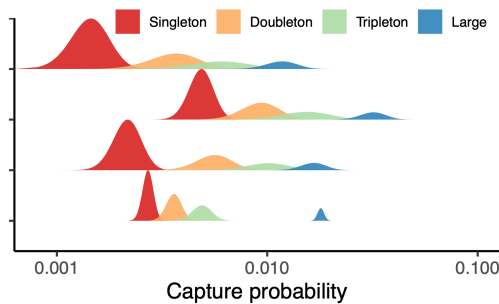
Observation: Clones of size s have a well-defined probabilities to be captured in the following sample that increase exponentially with s:
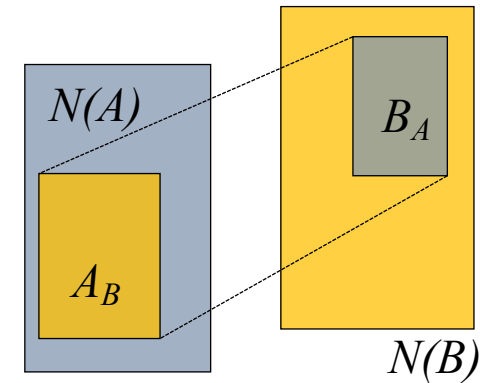
Solution: Group clones by size s and build a linear model (using common R-style syntax)

$$\log P \sim s + N$$

where P is the fraction of clones from A captured in B.
Condition/clone grouping effects are tested by introducing factor G

$$\log P \sim s + N + G$$

Applications: Allows quantifying
- chimerism/subset survival in HSCT

- clonal dynamics in time courses

- emerging antigen-specific T-cells