## Extra Slide on the Collision Algorithm (non-examinable)

- + The algorithm runs in (expected) sublinear time  $O(\sqrt{N})$ , where N := |S|
- The algorithm does not take a pre-specified and fixed number of samples

What can we do with a fixed number of samples *n*?

- We cannot find an unbiased estimator that works for any N (similar to Lecture 10, Slide 22)
- Could use hypothesis testing: For a fixed sample (x<sub>1</sub>, x<sub>2</sub>,..., x<sub>n</sub>) with c collisions, what is the probability to have c collisions under hypothesis that N ≥ x (or N = x) for some value x?
- Bayesian Approach: Assume unknown parameter *N* comes from a (known) probability distribution (called prior distribution). For a fixed sample (*x*<sub>1</sub>, *x*<sub>2</sub>,..., *x<sub>n</sub>*), update the probability distribution (called posterior distribution)

$$N \sim \operatorname{Exp}(1/1000) \xrightarrow{X_1 = x_1, \ldots, X_n = x_n} N \sim \left( \operatorname{Exp}(1/1000) \mid X_1 = x_1, \ldots, X_n = x_n \right)$$