Empirical reward vs true reward

- **Strong Law of large number:** Empirical means coverge to the true mean with probability 1. **But at what rate?**

- **Hoeffding's inequality :**

Theorem: Let $X_1, X_2, \cdots, X_n$ be n i.i.d. samples from a distribution over $[0,1]$ with mean μ, then

$$\mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) \leq 2\exp(-2n\epsilon^2).$$

Sample means in a typical set

- We define $\text{UCB}_a(t) = \hat{\mu}_a(t) + \sqrt{\dfrac{2\log(T)}{N_a(t)}}$ and $\text{LCB}_a(t) = \hat{\mu}_a(t) - \sqrt{\dfrac{2\log(T)}{N_a(t)}}$
- Using **Hoeffding's inequality**, we can prove:

$$\mathbb{P}\left(\text{for all arm } a \text{ and iteration } t,\ \text{LCB}_a(t) < \mu_a < \text{UCB}_a(t) \right) \geq 1 - \frac{1}{T}.$$

- Maximum regret suffered over T iterations is T.
- Regret suffered outside the good event is bounded by $\dfrac{1}{T} \times T = 1$, which

  contributes a constant. Hence we ignore everything outside the good event.
- Thus, we assume the sample means forever stay between the UCB and LCB

  for every arm.