Sample means in a typical set

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

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

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

Sample means in a typical set

- From now on, we restrict ourselves to the good event where the actual mean is contained in the interval $\left[\mathrm{LCB}_a(t),\ \mathrm{UCB}_a(t)\right]$
- **An interesting observation:** The interval $\left[\mathrm{LCB}_a(t),\ \mathrm{UCB}_a(t)\right]$ shrinks as we collect more samples from arm a.