Cumulated regrets for different bandit algorithms, averaged 1000 times
5 arms: $[B(0.6), B(0.2), B(0.55), B(0.7)^*, B(0.5)]$

Legend:
- Pure exploration
- Pure exploitation
- ε-greedy
- Explore-then-Exploit
- UCB($\alpha=1$)
- Bernoulli kl-UCB
- Thompson sampling
- [Lai & Robbins] lower bound $= 10.7 \log(t)$

Y-axis: Regret $R_t = t\mu^* - \sum_{s=1}^{t} \sum_{k=1}^{5} \mu_k \mathbb{E}_{1000}[T_k(t)]$

X-axis: Time steps $t = 1 \ldots T$, horizon $T = 10000$