

Bonus: Acquisition Function Comparison Study

This section compares multiple acquisition functions for Bayesian optimization.

Methodology

Acquisition Functions Compared

- **Random Search** (baseline)
- **EI ($\xi=0.01$)**: Expected Improvement with exploration margin
- **PI ($\xi=0.01$)**: Probability of Improvement with exploration margin
- **LCB ($\kappa=1$)**: Lower Confidence Bound with low exploration
- **LCB ($\kappa=2$)**: Lower Confidence Bound with moderate exploration

Experimental Setup

Parameter	Value
Initial observations	5 (random, shared across all methods)
BO iterations	30
Total evaluations	35
GP Model	RBF kernel with log(y+1) transform
Number of runs	20 (paired comparison)

IMPORTANT: All acquisition functions share identical initial points per run for proper paired comparison. This was verified with runtime assertion.

Results

Branin Function

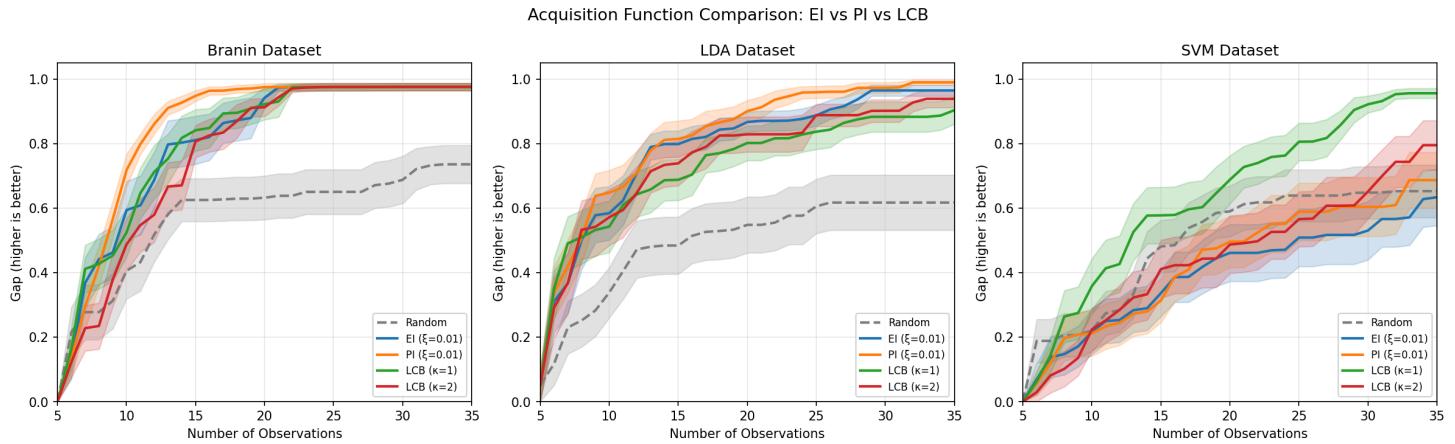


Figure 1: Learning curves with \pm SE bands (x-axis starts at 5).

Rankings (n=20 runs)

Rank	Method	Mean Gap	\pm SE
1	LCB ($\kappa=1$)	0.974	± 0.011
2	PI ($\xi=0.01$)	0.974	± 0.011
3	EI ($\xi=0.01$)	0.969	± 0.013
4	LCB ($\kappa=2$)	0.963	± 0.013
5	Random	0.736	± 0.059

Paired t-tests (LCB $\kappa=1$ vs others):

- vs PI: $p=0.55$, $d=+0.00$ (n.s.) — **statistically indistinguishable**
- vs EI: $p=0.14$, $d=+0.09$ (n.s.)
- vs Random: $p=0.0004$, $d=+1.26^*$ — **large effect size**

LDA Dataset

Rank	Method	Mean Gap	\pm SE
1	PI ($\xi=0.01$)	0.948	± 0.027

Rank	Method	Mean Gap	\pm SE
2	LCB ($\kappa=1$)	0.937	± 0.023
3	LCB ($\kappa=2$)	0.891	± 0.034
4	EI ($\xi=0.01$)	0.876	± 0.056
5	Random	0.617	± 0.085

Paired t-tests (PI vs others):

- vs LCB ($\kappa=1$): $p=0.78$, $d=+0.09$ (n.s.)
- vs Random: $p=0.001$, $d=+1.17^*$

SVM Dataset

Rank	Method	Mean Gap	\pm SE
1	LCB ($\kappa=1$)	0.839	± 0.055
2	Random	0.652	± 0.082
3	PI ($\xi=0.01$)	0.628	± 0.078
4	EI ($\xi=0.01$)	0.593	± 0.082
5	LCB ($\kappa=2$)	0.559	± 0.082

κ Sensitivity Analysis

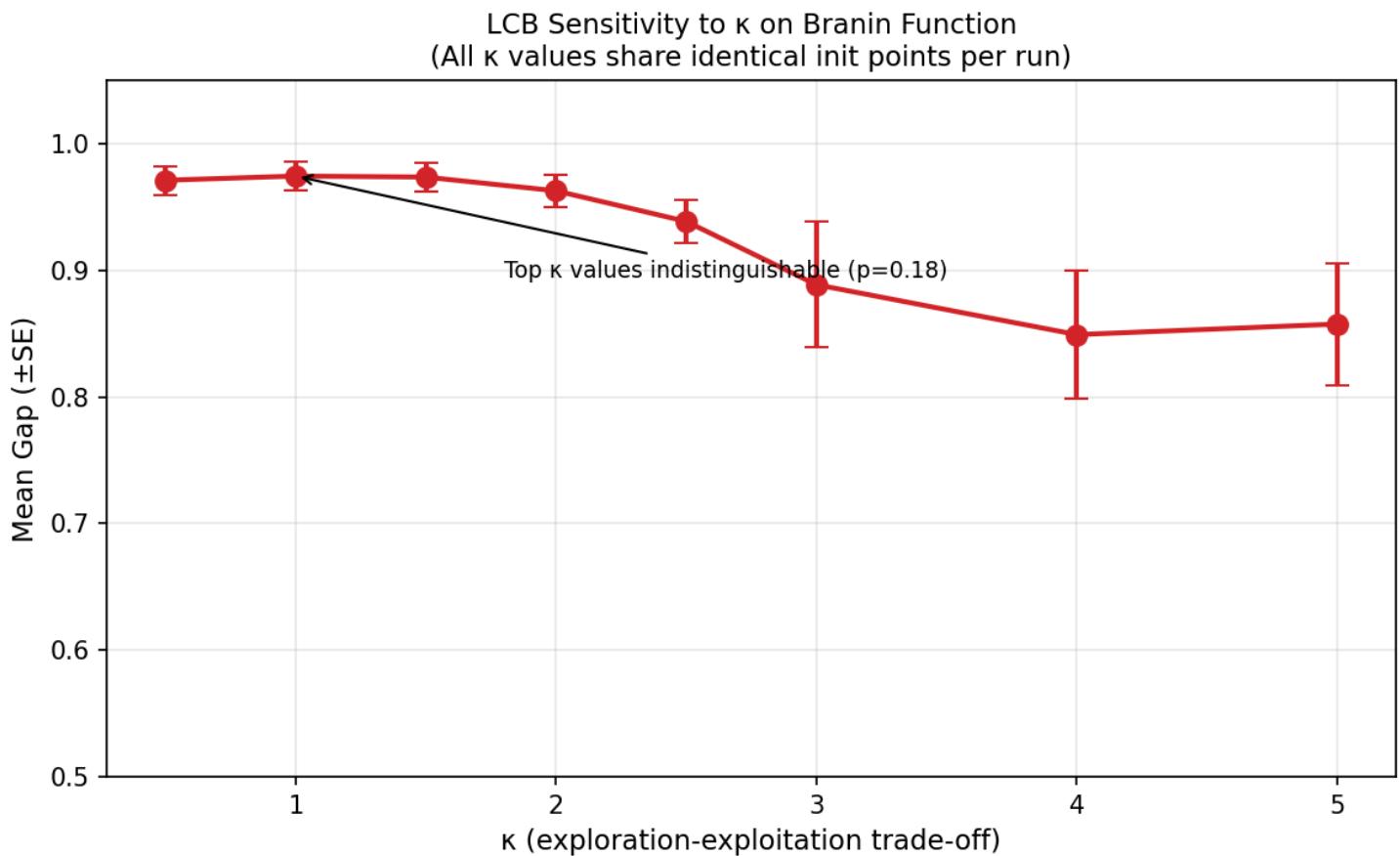


Figure 2: LCB sensitivity to κ on Branin (all κ values share identical init points per run).

Key Findings

1. All BO methods significantly outperform Random Search on Branin and LDA ($d > 1.0$)
2. Top acquisition functions are statistically indistinguishable on Branin ($LCB \kappa=1 \approx PI \approx EI$)
3. LCB ($\kappa=1$) performs well across all datasets — low exploration parameter works well here
4. SVM shows unexpected results — LCB ($\kappa=1$) best, but EI/PI underperform Random

Caveats

- $p > 0.05$ means "no significant difference detected," NOT "equivalence"
- Multiple comparisons not corrected (interpret p-values with caution)
- These results depend on the specific kernel choice (RBF) and log transform