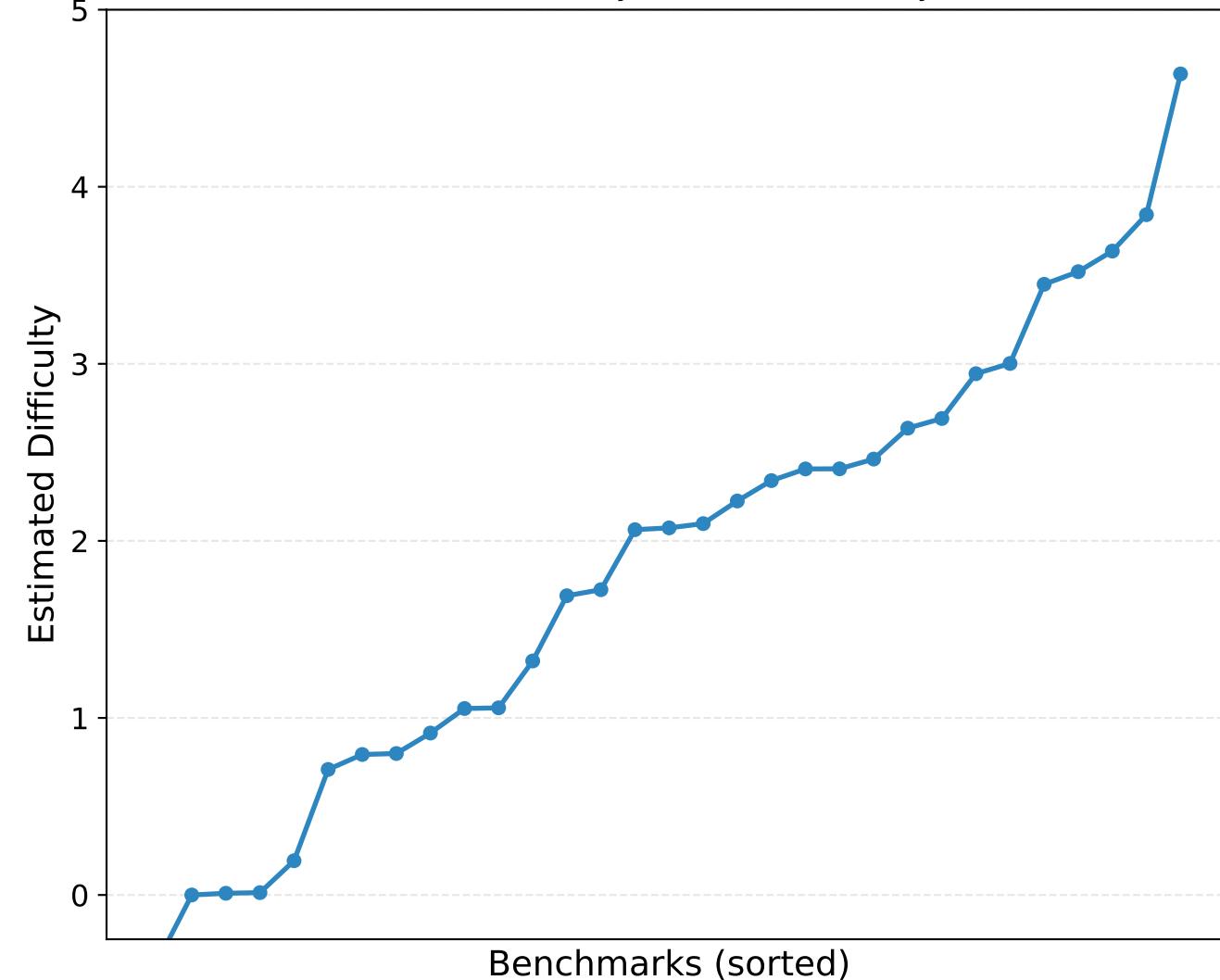


Benchmarks by Estimated Difficulty



Models by Estimated Capability

