Figure columns (left to right): pi0 = pythia1b, pi0 = pythia400m, pi0 = fixedrdm; second row: pi0 = pythia1b, pi0 = pythia400m, pi0 = fixedrdm; third row: pi0 = pythia7b, pi0 = pythia1b, pi0 = pythia400m.

Row labels (top to bottom): Code generation, Code critique, MATH.

X-axis: Generated samples. Y-axis: Correctness.

Legend:
- Locked with pwd
- Locked without pwd
- Expert iteration
- Filtering
- Iterated DPO
- DPO