

Test loss stdev

- Jorgensen mean threshold
- Jorgensen max threshold
- Weighted (ours)

0.1 0.2 0.3 0.4 0.5 0.6

Fraction of  
conservative users  $f_c$

