

График №4: Зависимость дисперсии градиентов от параметра  $\delta$  в AdaHeavy  
(28% шума, распределение Вейбулла,  $k=0.28$ , 20 запусков)  
→  $0.0 \pm 0.0 \times$  снижение дисперсии против Adam при  $\delta=0.10$  ( $p=0.9984$ )  
См. Теорему 3.4: Оптимальный выбор  $\delta$  для минимизации дисперсии

