

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

