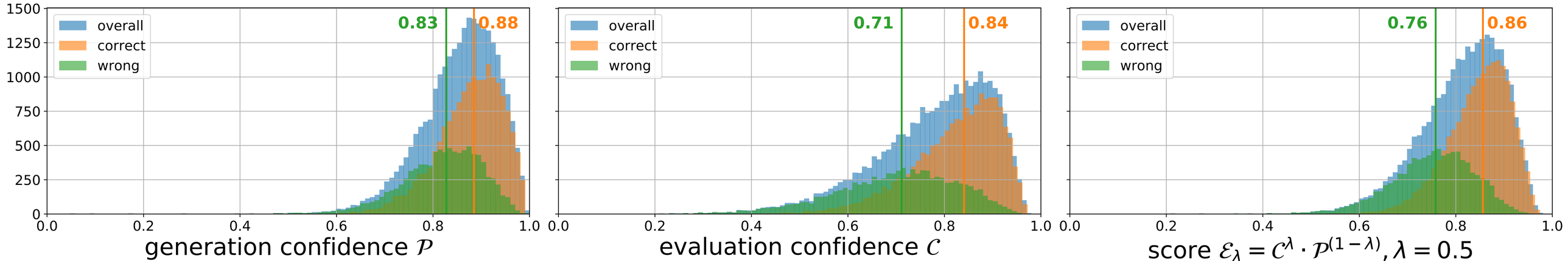
overall
correct
wrong
0.83
0.88
generation confidence $\mathcal{P}$
0.71
0.84
evaluation confidence $\mathcal{C}$
0.76
0.86
score $\mathcal{E}_\lambda = \mathcal{C}^\lambda \cdot \mathcal{P}^{(1-\lambda)}, \lambda = 0.5$